

An invariance principle for branching diffusions in bounded domains

Ellen Powell

Abstract

We study branching diffusions in a bounded domain D of \mathbb{R}^d in which particles are killed upon hitting the boundary ∂D . It is known that any such process undergoes a phase transition when the branching rate β exceeds a critical value: a multiple of the first eigenvalue of the generator of the diffusion. We investigate the system at criticality, and prove an asymptotic for the probability of survival up to large times. We show further that the genealogical tree associated with such a critical process converges to Aldous' Continuum Random Tree under appropriate rescaling. The result holds under only a mild assumption on the domain, and is valid for all branching mechanisms with finite variance, and a general class of diffusions.

1 Introduction

This paper concerns branching diffusions in a bounded domain D of \mathbb{R}^d . These are processes in which individual particles move according to the law of some diffusion, are killed upon exiting the domain, and branch into a random number of particles (with distribution A , independent of position) at rate $\beta > 0$. Whenever such a branching event occurs, each of the offspring then independently repeats the behaviour of its parent, starting from the point of fission. Throughout, the configuration of particles will be denoted by

$$(X_t^1, \dots, X_t^{N_t})$$

where N_t is the number of particles alive at time t , and we will write \mathbb{P}_x for the law of the process initiated from a point $x \in D$. We will always assume that the offspring distribution has mean $m > 1$ and finite variance, and that the generator $L = -\frac{1}{2} \sum_{i,j} a^{ij} \partial x_i \partial x_j + \sum_i b^i \partial x_i$ of the diffusion is uniformly elliptic and self-adjoint with smooth coefficients.

It is known that such a system exhibits a phase transition in the branching rate: for large enough β there is a positive probability of survival, but for small β , including at criticality, there is almost sure extinction. The critical value of β is equal to $\frac{\lambda}{m-1}$, where λ is the first eigenvalue of L on D with Dirichlet boundary conditions. The main goal of this paper will be to study the system at criticality and find a scaling limit for the resulting genealogical tree. This is the continuous planar tree that is generated purely by the birth and death times of particles in the system, and encodes no information about the spatial movement. More precisely, for given $y > 0$, we condition the diffusion to survive until time $ny > 0$ and look at the associated genealogical tree \mathcal{T}_n^y , equipped with its natural distance d_n^y . Rescaling the distances by a factor n produces a sequence of random compact metric spaces $(\mathcal{T}_n^y, \frac{1}{n} d_n^y)_{n \in \mathbb{N}}$. We will prove that this sequence converges in distribution to a conditioned Brownian Continuum Random Tree as $n \rightarrow \infty$, with respect to the Gromov-Hausdorff topology. Indeed, if we write $(\mathcal{T}_{e^y}, d_{e^y})$ for the real tree whose contour function is given by e^y , a Brownian excursion conditioned to reach level y , then we obtain the following result.

Theorem 1.1. *Suppose that $D \subset \mathbb{R}^d$ is a C^1 domain and that L is uniformly elliptic and self-adjoint with smooth coefficients. Further suppose that A has mean $m > 1$ and finite variance, and $\varphi \in C^1(\overline{D})$ where φ is the first eigenfunction of L on D . Then for any $y > 0$, and any starting point $x \in D$,*

$$(\mathcal{T}_n^{\alpha y}, \frac{1}{\alpha n} d_n^{\alpha y}) \xrightarrow{n \rightarrow \infty} (\mathcal{T}_{e^y}, d_{e^y})$$

in distribution, with respect to the Gromov-Hausdorff distance, where

$$\alpha = \sqrt{\frac{4(m-1)}{\lambda(1, \varphi) \mathbb{E}[A^2 - A] \int_D \varphi(y)^3 dy}}.$$

Remark 1.2. *One sufficient condition to ensure that the hypotheses of Theorem 1.1 are satisfied is to assume that the boundary of D is $C^{2+[\frac{d}{2}]}$. Standard regularity theory of elliptic partial differential equations, see for example [Eva98, §6.3], then implies that $\varphi \in C^1(\overline{D})$. However, this is also satisfied in many other cases.*

On the way to proving Theorem 1.1 we obtain several other results on critical branching diffusions, which are interesting in their own right as well as being essential to our method. We start with the phase transition. This was first proved by Sevast'yanov [Sev58] and Watanabe [Wat65], but has also been reworked and generalised in recent years, for example in [EK04], which studies local versus global extinction in unbounded domains. The precise description is as follows:

Theorem 1.3 ([Sev58], [Wat65]). *Let $D \subset \mathbb{R}^d$ be a bounded domain, satisfying a minimal regularity assumption (see Condition 2.1). Suppose that L is a uniformly elliptic self-adjoint operator with smooth coefficients and that A is a distribution with finite mean $m > 1$. Then there are two possibilities for the long term behaviour of the branching diffusion determined by L and A in D , according to the value of the branching rate β . Namely, for any starting position $x \in D$, if λ is the principal eigenvalue of L on D with Dirichlet boundary conditions then,*

(1) *for $\beta > \frac{\lambda}{m-1}$ the process survives for all time with positive probability.*

(2) *for $\beta \leq \frac{\lambda}{m-1}$ the process becomes extinct almost surely.*

Moreover, if $\beta \leq \frac{\lambda}{m-1}$ then $\mathbb{P}_x(N_t > 0) \rightarrow 0$ uniformly in D .

In the statement above we have taken some care to specify the regularity required on the domain, which is not detailed in the earlier works. Essentially we require that the eigenfunctions of the Laplacian converge to 0 pointwise on the boundary, and that Brownian motion started from a point on the boundary leaves the domain immediately with probability one. We will provide alternative proof of this Theorem, which in contrast to the earlier more analytic proofs in [Sev58], [Wat65], uses arguments centred around martingales (also appearing in [EK04]) arising naturally from the definition of the process. This proof will show that the stated regularity assumptions are sufficient.

The rest of the paper will focus on the behaviour of the system at criticality, starting with an asymptotic for the survival probability.

Theorem 1.4. *Suppose that the domain D is C^1 , that L is as in Theorem 1.3, and that A has finite variance. Then, in the critical case $\beta = \frac{\lambda}{m-1}$, for all $x \in D$ we have*

$$\mathbb{P}_x(N_t > 0) \sim \frac{1}{t} \times \frac{2(m-1)\varphi(x)}{\lambda(\mathbb{E}[A^2] - \mathbb{E}[A]) \int_D \varphi(y)^3 dy} \quad (1.1)$$

as $t \rightarrow \infty$. Here φ is the first eigenfunction of L on D , normalised to have unit L^2 norm.

This asymptotic then allows us to study the behaviour of the system when it is conditioned to survive for a long time, which is important for the proof of Theorem 1.1. One tool that we will use is a classical *spine* change of measure, under which the process has a distinguished particle, the spine, which is conditioned to remain in D forever (as in [Pin85]). Along this spine, families of ordinary critical branching diffusions immigrate at rate $\frac{m}{m-1}\lambda$ according to a biased offspring distribution. Note that there is no extinction under this new measure, which we denote by \mathbb{Q}_x . We will prove that changing measure in this way is in fact somewhat close to conditioning on survival for all time, in the sense of the following Proposition.

Proposition 1.5. *Assume the hypotheses of Theorem 1.4. Then for any $T \geq 0$, $x \in D$ and $B \in \mathcal{F}_T$, where \mathcal{F} is the natural filtration of the process, we have*

$$\lim_{t \rightarrow \infty} \mathbb{P}_x(B|N_t > 0) = \mathbb{Q}_x(B). \quad (1.2)$$

Furthermore, we are able to prove a Yaglom type limit theorem for the positions of the particles in the system at time t , given survival.

Theorem 1.6. *For any measurable function f on D such that $\int_D f(x)^2 \varphi(x) dx < \infty$, we have*

$$\left(t^{-1} \sum_{i=1}^{N_t} f(X_t^i) \middle| N_t > 0 \right) \rightarrow Z$$

in distribution as $t \rightarrow \infty$, where Z is an exponential random variable with mean

$$\frac{\lambda (\mathbb{E}[A^2] - \mathbb{E}[A]) (\varphi, f)_{L^2(D)} \int_D \varphi^3}{2(m-1)}.$$

One consequence of Theorem 1.6 (or rather its proof) is that it allows us to describe the limiting distribution of the particles in the system at time t , given survival. It turns out that this is the law with density φ , normalised to be a probability distribution.

Corollary 1.7. *Let*

$$\mu_t := \frac{1}{N_t} \sum_{i=1}^{N_t} \delta_{X_t^i}$$

be the uniform distribution on all particles alive at time t , given survival. Then, for each f as in Theorem 1.6, we have that

$$\mu_t(f) \rightarrow \mu(f)$$

in distribution, and hence in probability, as $t \rightarrow \infty$, where

$$\mu(f) = \frac{\int_D \varphi(x) f(x) dx}{\int_D \varphi(x) dx}.$$

As this paper was being completed, the author learnt that similar results to Theorem 1.4 and Theorem 1.6 have also been shown by Asmussen and Hering in [AH83]. However their proofs are completely different from those in the current paper and, more importantly, our method provides several new and crucial ingredients for the proof of Theorem 1.1.

1.1 Related Work

It is interesting to note the analogy between Theorems 1.3-1.6, and the classical results from the theory of Galton-Watson processes. Indeed, for critical Galton-Watson processes, Kolmogorov [Kol38] proved an asymptotic for the probability of survival up to time n ;

$$\mathbb{P}(Z_n > 0) \sim \frac{c}{n}$$

where Z_n is the population size at time n , and the constant depends on the variance of the offspring distribution. Moreover, Aldous [Ald91],[Ald93] and Duquesne and Le Gall [GD02] showed that if you condition a critical Galton-Watson process to reach a large generation or have a large total progeny, then you have a scaling limit for the resulting tree. This limit is in the Gromov-Hausdorff topology, after rescaling distances in the tree appropriately, and the limiting object is the Continuum Random Tree, [Ald91]. In fact, this result can be extended to multitype Galton-Watson processes, as in [Mie08], where the same scaling limit exists. Since a branching diffusion can be thought of as a limit of multitype Galton-Watson processes, considering the types to be positions and discretising the domain appropriately, this is the first indication we should be able to obtain a similar scaling limit.

Remark 1.8. *The constant α in Theorem 1.1 is exactly what one obtains formally from the convergence in [Mie08], considering the branching diffusion to be a scaling limit of appropriate multitype Galton-Watson processes. However, Miermont's proof strategy is to make an induction on the number of types, and the lack of uniformity in the estimates as the number of types grows means that it does not extend to the set up considered here. Instead we use a combination of probabilistic and analytic ideas; see Section 1.2 for a sketch of the argument.*

An asymptotic for the survival probability has also been considered previously, see [BBS14] and [Kes78], in the case of branching Brownian motion with absorption at the origin, where the branching rate is kept constant and each particle moves with a drift $-\mu$, which is varied. In this set up, there is a critical value of $\mu = \mu_c$ above which extinction occurs with probability one. The *near-critical* system, as μ approaches its critical value from below, has also been studied, and in [BBS11] a limit, as $\mu \uparrow \mu_c$, is found for the probability of survival for all time as a function of the initial position. However, these results are quite different from ours as we do not allow our domain to be unbounded. The proofs of Theorems 1.1 and 1.4 do not extend to this situation, and in fact, we would expect to see a variety of behaviours for the critical system if we remove the assumption of boundedness. It would be an interesting problem to explore the possible cases here, and classify which domains fall into the regime of Theorem 1.1 and Theorem 1.4.

1.2 Organisation of the Paper and Main Ideas

We begin, for completeness and in order to introduce key concepts for the latter part of the paper, by providing a full proof of Theorem 1.3 in the case of Brownian motion with binary branching. This also allows us to make precise the regularity that is required on the domain for this statement to be true, see Condition 2.1. The main idea behind the proof we will give is to exploit the existence of a certain martingale

$$M_t = e^{(\lambda-\beta)t} \sum_{i=1}^{N_t} \varphi(X_t^i),$$

where φ is the first eigenfunction of $-\frac{1}{2}\Delta$ on D , with unit L^2 norm, and λ is the first eigenvalue. We show that its properties (which depend on β) change critically at the point $\beta = \lambda$. These critical features turn out to determine the long term behaviour of the entire process, and thus provide the result of the theorem.

We will then turn to the proofs of the remaining theorems. Again we will give these for binary branching Brownian motion, and wherever adaptation for general diffusions and branching mechanisms is required, we will indicate the necessary changes. Any extra arguments are in fact minor, which is why we prefer to highlight the simplest case. This allows us to keep the arguments clear and avoid introducing extra notation.

The proof of Theorem 1.4 proceeds by a combination of probabilistic arguments, and analysis of the Fisher-Kolmogorov-Petrovskii-Piscounov equation. This is a partial differential equation which is known to be satisfied by the survival probability for branching Brownian motion; first noted by McKean [McK75] in the one-dimensional case, and used as the main tool by Sevast'yanov in the original proof of Theorem 1.3. Naively, we can write the survival probability in an L^2 expansion with respect to the orthonormal basis of $L^2(D)$ given by eigenfunctions of the Laplacian. Since the survival probability satisfies the FKPP equation we get a family of coupled ODEs from the coefficients. However, this is tricky to analyse directly. Instead, we apply a probabilistic line of reasoning, changing measure by M_t/M_0 to get a spine characterisation of the system as discussed in the introduction. This allows us to deduce that the survival probability decays like $a(t)\varphi(x)$ as $t \rightarrow \infty$, where $a(t)$ is the first coefficient in our expansion. Thus, our problem is reduced to the study of a single ODE. From here elementary analysis, combined with some extra information obtained from the probabilistic arguments, yields the result. We then prove Theorem 1.6 and Corollary 1.7, using the method of moments and a *Many-to-Few* Lemma.

The remainder of the paper is devoted to the proof of Theorem 1.1. Again working primarily in the case of Brownian motion with binary branching, we take an i.i.d. sequence of critical processes and concatenate the height functions of their associated trees. We would like to find a process which approximates this, and will converge after rescaling to a reflected Brownian motion: an analogue of the *Lukasiewicz path* for Galton-Watson trees. Just as the

martingale M_t roughly measures the size of our system as we increase time, exploring it in a different, *depth-first*, order provides another martingale that is a proxy for the height function. After strengthening our result, Corollary 1.7, for the conditioned system, we can prove that the quadratic variation of this martingale is essentially linear, and so we obtain an invariance principle.

We then have to prove that this martingale is indeed a good approximation to the height process. This is one of the main difficulties, as the reversibility tools that are key to proving this for the Lukasiewicz path in the Galton-Watson case are lost. Instead, we must use precise estimates, and a delicate ergodicity argument related to our spine change of measure. This is one of the reasons that our machinery from the proof of Theorem 1.4 is so essential. Tightness arguments then allow us to conclude.

Acknowledgements I would like to thank Nathanaël Berestycki, for suggesting this problem and for many helpful discussions and suggestions.

2 Preliminaries

2.1 Diffusions as Trees

As stated in the introduction, \mathbb{P}_x will denote the law of our branching diffusion, initiated from the point $x \in D$. We will consider this as a law on continuous planar trees, where every vertex is also marked with a position in D . In this representation, the vertices at a given height t in the tree will correspond to the particles alive in the system at time t , and their marks will correspond to their positions. To complete the definition of the tree we also need to decide, at every branch point, how the branching subtrees are ordered from left to right. To do this, we assume that given the number of offspring at a branching event, this ordering is chosen uniformly at random. This gives us a law on planar, or equivalently labelled, rooted trees. We emphasise here that these trees, unlike Galton-Watson trees, are in *continuous time*. Note that, in Theorem 1.1, we are considering them without their marks (in fact, the marks are irrelevant to the tree considered as a metric space.) The final point to make, is that when we use our notation $(X_t^1, \dots, X_t^{N_t})$ for the system at time t , the indices correspond to the ordering of the vertices from left to right in the tree, and the X_t^i are their positions, or marks.

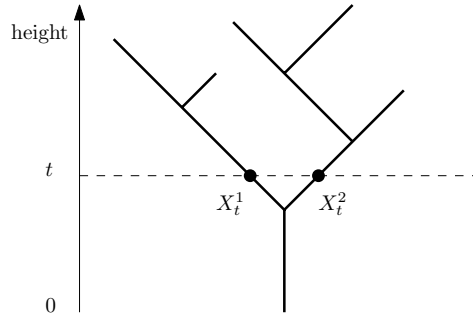


Figure 2.1: An example of the continuous tree generated by a branching diffusion. The two vertices at height t (marked with dots) have positions, or marks, given by X_t^1 and X_t^2 .

2.2 Spectral Theory and Martingales

From now on we will work in the case of Brownian motion with binary branching, unless stated otherwise.

As discussed in the introduction, the behaviour of branching Brownian motion killed when leaving a given domain D will be closely related to the spectral properties of the Laplacian on that domain. Throughout, we will let $\{\lambda_i\}_{i \geq 1}$ denote the eigenvalues of $-\frac{1}{2}\Delta$ on D with Dirichlet boundary conditions, with corresponding eigenfunctions $\{\varphi_i\}_{i \geq 1}$, normalised to have unit L^2 norm. Recall that under this normalisation the eigenfunctions

form an orthonormal basis for $L^2(D)$, the eigenvalues are real with

$$\lambda := \lambda_1 < \lambda_2 \leq \lambda_3 \leq \dots \rightarrow \infty$$

and the first eigenfunction

$$\varphi := \varphi_1$$

is strictly positive in the domain. Also note that the eigenfunctions are in $C^\infty(D)$, and assuming the domain is sufficiently regular, converge pointwise to 0 on ∂D . In particular, they are all bounded.

Condition 2.1. *For the proof of Theorem 1.3 we will assume that the domain D is regular enough that:*

(1) *The eigenfunctions converge to 0 pointwise on the boundary, and*

(2) *Every point x_0 on the boundary satisfies, for every $\varepsilon > 0$, $\lim_{x \rightarrow x_0} \mathbf{P}_x(\tau_D > \varepsilon) = 0$,*

where \mathbf{P}_x is the law of Brownian motion started from the point x and τ_D is the first time it hits the boundary ∂D .

These are very weak regularity conditions on the domain D . For example, see [GT83, Theorem 8.29], both conditions are satisfied by any domain satisfying a uniform exterior cone condition.

To set up some notation, let $p^{\mathbb{R}^d}(t, x, y)$ be the transition density for Brownian motion on \mathbb{R}^d and set, for $x, y \in D$,

$$p^D(t, x, y) = p^{\mathbb{R}^d}(t, x, y) - \mathbf{E}_x \left[p^{\mathbb{R}^d}(t - \tau_D, B_{\tau_D}, y) \mathbb{1}_{\{\tau_D \leq t\}} \right].$$

Then we have

$$0 \leq p^D(t, x, y) \leq p^{\mathbb{R}^d}(t, x, y) \leq \frac{1}{(2\pi t)^{d/2}} \quad (2.1)$$

and by the strong Markov property p^D is the transition density of Brownian motion *killed* when leaving the domain D . This means, in particular, that for all integrable functions f we have

$$\int_D f(y) p^D(t, x, y) dy = \mathbf{E}_x [f(B_t) \mathbb{1}_{\{\tau_D > t\}}] \quad (2.2)$$

for $t > 0$ and $x \in D$. It can be shown (see for example [MNV09, Remark 2.1]) that if D satisfies Condition 2.1, then for any bounded continuous function f on D which vanishes on ∂D ,

$$\int_D p^D(t, x, y) f(y) dy = \mathbf{E}_x [f(B_t) \mathbb{1}_{\{\tau_D > t\}}] \quad (2.3)$$

is the unique solution of the heat equation in the domain, with initial data f and Dirichlet boundary data. One consequence of this is that for any of the eigenfunctions φ_i , we have

$$\mathbf{E}_x [\varphi_i(B_{t \wedge \tau_D})] = e^{-\lambda_i t} \varphi_i(x) \quad (2.4)$$

for all $x \in D$. This leads to the following decomposition for functions $f \in L^2(D)$. Here and throughout the rest of the paper, (\cdot, \cdot) will represent the usual inner product on $L^2(D)$.

Lemma 2.2. *If $f \in L^2(D)$ then for all $t \geq 0$ and $x \in D$*

$$\mathbf{E}_x [f(B_t) \mathbb{1}_{\{\tau_D > t\}}] = \sum_{i=1}^{\infty} e^{-\lambda_i t} \varphi_i(x) (\varphi_i, f).$$

Proof. First note that $p^D(t, x, y)$ is bounded for all $x, y \in D$ by (2.1), and that $(p^D(t, x, \cdot), \varphi_i) = e^{-\lambda_i t} \varphi_i(x)$ by (2.3) and (2.4). Therefore, since the φ_i 's form an orthonormal basis of $L^2(D)$, we have that

$$\sum_{i=1}^n e^{-\lambda_i t} \varphi_i(x) \varphi_i(\cdot) \rightarrow p^D(t, x, \cdot)$$

in L^2 . Thus for any $f \in L^2(D)$, we may conclude that

$$\mathbf{E}_x [f(B_t) \mathbb{1}_{\{\tau_D > t\}}] = (p^D(t, x, \cdot), f) = \lim_{n \rightarrow \infty} \left(\sum_{i=1}^n e^{-\lambda_i t} \varphi_i(x) \varphi_i(\cdot), f \right)$$

which yields the result. \square

Another consequence of (2.4) is the existence of an additive martingale for the system. Indeed, a straightforward application of the branching Markov property and the Many-to-One Lemma (see for example [HH07] or [HR15]) which tells us that

$$\mathbf{E}_x \left[\sum_{i=1}^{N_t} f(X_t^i) \right] = e^{\beta t} \mathbf{E}_x [f(B_t) \mathbb{1}_{\{\tau_D > t\}}]$$

for measurable f , provides the following:

Lemma 2.3. *The process*

$$M_t = e^{(\lambda - \beta)t} \sum_{i=1}^{N_t} \varphi(X_t^i)$$

is a martingale under \mathbb{E}_x , for each $x \in D$.

Moreover, by positivity of φ in D , M_t is a positive martingale, and thus by the martingale convergence theorem converges almost surely to an almost surely finite limit. This is a strong indicator of the existence of a phase transition, since the properties of the limit M_∞ , which depend on $\beta > 0$ and change critically at $\beta = \lambda$, determine the long term behaviour of the system.

Remark 2.4. *The above theory directly extends to more general diffusions, with generators and offspring distributions as in Theorem 1.3. In particular, Condition 2.1 will provide the required degree of regularity. Here we have that*

$$M_t = e^{(\lambda - \beta(m-1))t} \sum_{i=1}^{N_t} \varphi(X_t^i)$$

defines a positive martingale, using a generalisation of the Many-to-One Lemma.

3 The Phase Transition

In this section we will provide a proof of Theorem 1.3.

3.1 The Supercritical Case

We begin by supposing that β is strictly greater than λ , in which case the proof can be summarised as follows. We already know that $M_t(\beta) \rightarrow M_\infty$ as $t \rightarrow \infty$. We will show that for this range of β , the martingale M_t is square integrable and so the limit M_∞ cannot be degenerate. Thus, since $e^{(\lambda - \beta)t} \rightarrow 0$ as $t \rightarrow \infty$, it must be the case that

$$\sum_{i=1}^{N_t} \varphi(X_t^i) \rightarrow \infty$$

with positive probability. Boundedness of φ then implies that $N_t \rightarrow \infty$ on this event.

In order to prove square integrability, we need a generalisation of the Many-to-One Lemma to a Many-to-Two form. This is well known, see for example [HR15, Lemma 1] or [Wat67, Equation (2.23)].

Lemma 3.1 (Many-to-Two). *If f, g are measurable functions on D then*

$$\mathbb{E}_x \left[\sum_{i=1}^{N_t} f(X_t^i) \sum_{i=1}^{N_t} g(X_t^i) \right] = \mathbf{E}_x \left[e^{\beta t} f(B_t) g(B_t) \mathbb{1}_{\{\tau_D > t\}} \right] + \mathbf{E}_x \left[\int_0^{t \wedge \tau_D} 2\beta e^{\beta s} \mathbf{E}_{B_s} \left[\sum_1^{N_{t-s}} f(X_{t-s}^i) \right] \mathbf{E}_{B_s} \left[\sum_1^{N_{t-s}} g(X_{t-s}^i) \right] ds \right]$$

where B is standard Brownian motion started at $x \in D$ and $(X_t^1, \dots, X_t^{N_t})$ represents branching Brownian motion as usual.

Applying the Lemma with $f = g = \varphi$, and using (2.4) to rewrite the expectation terms in the integrand, along with Fubini, we see that

$$\mathbb{E}_x [M_t^2] = e^{2(\lambda-\beta)t} \mathbf{E}_x \left[e^{\beta t} \varphi(B_{t \wedge \tau_D})^2 \right] + 2\beta \int_0^t e^{2(\lambda-\beta)s} \mathbf{E}_x \left[e^{\beta s} \varphi(B_{s \wedge \tau_D})^2 \right] ds.$$

Since we also have the bound,

$$\mathbf{E}_x \left[e^{\beta t} \varphi(B_{t \wedge \tau_D})^2 \right] \leq \|\varphi\|_\infty \mathbf{E}_x \left[e^{\beta t} \varphi(B_{t \wedge \tau_D}) \right] \leq \varphi(x) \|\varphi\|_\infty e^{(\beta-\lambda)t}$$

for all t and x , we can substitute this in and integrate to see that

$$\mathbb{E}_x [M_t^2] \leq \varphi(x) \|\varphi\|_\infty \left(1 + \frac{4\beta}{\lambda - \beta} \right)$$

for all t . Thus we obtain square integrability.

Remark 3.2. *For the more general set up, with generator L and offspring distribution A as in Theorem 1.3, a generalisation of the Many-to-Two lemma, see [HR15], proves uniform integrability of the martingales. This again shows that there is a positive probability of survival in the supercritical case.*

3.1.1 The Subcritical Case

Now let us suppose that $\beta < \lambda$. The convergence of the martingale in this case, along with the fact that $e^{(\lambda-\beta)t} \rightarrow \infty$ as $t \rightarrow \infty$, means that in fact

$$\sum_{i=1}^{N_t} \varphi(X_t^i) \rightarrow 0$$

almost surely as $t \rightarrow \infty$. Although this does not show that $N_t \rightarrow 0$ a.s. for all such β immediately, Lemma 2.2 ensures that $\sum_{i=1}^{N_t} \varphi(X_t^i)$ being small must mean that N_t is small with high probability. This is made explicit by the following Lemma, which is essentially a consequence of Lemma 2.2.

Lemma 3.3. *Let $\varepsilon > 0$ be given. Then there exists $T(\varepsilon) \geq 0$ and an absolute constant K , such that for all $t \geq T(\varepsilon)$ and all $x \in D$ we have*

$$\mathbf{P}_x (\{\varphi(B_t) < \varepsilon\} \cap \{t < \tau_D\}) \leq e^{-\lambda t} K \varepsilon.$$

Proof. Let $f_\varepsilon(x) = \mathbb{1}_{[0, \varepsilon)}(\varphi(x)) = \mathbb{1}_{[0, \varepsilon)} \cdot \varphi$. This is clearly in L^2 (it is bounded by 1), so we may apply Lemma 2.2 to obtain that

$$\mathbf{P}_x (\{\varphi(B_t) < \varepsilon\} \cap \{t < \tau_D\}) = \mathbf{E}_x [f_\varepsilon(B_t) \mathbb{1}_{\{\tau_D > t\}}] = \sum_{i=1}^{\infty} e^{-\lambda_i t} (\varphi_i, f_\varepsilon) \varphi_i(x).$$

Intuitively, since $\lambda = \lambda_1 < \lambda_i$ for $i \geq 2$, we expect that the sum should behave roughly like $e^{-\lambda t} (\varphi, f_\varepsilon) \varphi(x)$ as t becomes large. Indeed we can bound it above, using that $(\varphi, f_\varepsilon) \leq \text{vol}(D)\varepsilon$, by

$$e^{-\lambda t} \left(\|\varphi\|_\infty \text{vol}(D)\varepsilon + e^{-\gamma t} \left| \sum_{i=2}^{\infty} e^{-(\lambda_i - \lambda_2)t} (\varphi_i, f_\varepsilon) \varphi_i(x) \right| \right) \quad (3.1)$$

where $\gamma > 0$ is the spectral gap for D . Therefore, it is enough to show that the expression in the modulus above is bounded by some absolute constant, for all t large enough. To do this, observe that since $\lambda_i \rightarrow \infty$ as $i \rightarrow \infty$, there exists an N such that $(\lambda_i - \lambda_2) > \frac{\lambda_i}{2}$ for all $i \geq N$. This means that

$$\begin{aligned} \left| \sum_{i=2}^{\infty} e^{-(\lambda_i - \lambda_2)t} (\varphi_i, f) \varphi_i(x) \right| &\leq \sum_{i=2}^{\infty} e^{-(\lambda_i - \lambda_2)t} |(\varphi_i, f)| |\varphi_i(x)| \\ &\leq N \text{vol}(D) \sup_{1 \leq i \leq N} \|\varphi_i\|^2 + \sum_{N+1}^{\infty} e^{-\frac{\lambda_i}{2}t} |(\varphi_i, f)| |\varphi_i(x)| \end{aligned}$$

for all t , where the first term in the final line is a constant depending only on D . Furthermore, by Cauchy-Schwarz we have

$$\left| \sum_{N+1}^{\infty} e^{-\frac{\lambda_i}{2}t} |(\varphi_i, f)| |\varphi_i(x)| \right| \leq \sqrt{\sum_{i=N+1}^{\infty} e^{-\lambda_i t} \varphi_i(x)^2} \sqrt{\sum_{i=N+1}^{\infty} (\varphi_i, f)^2}$$

where the first term is less than $\|p^D(t/2, x, \cdot)\|_{L^2} \leq (\pi t)^{-d/4}$ and the second is less than $\|f\|_\infty$. This can clearly be bounded uniformly for all t large enough. \square

Corollary 3.4. *Let τ_D be the hitting time of ∂D , for a Brownian motion started at $x \in D$. Then*

$$\mathbf{P}_x(\tau_D > t) \leq A e^{-\lambda t}$$

for all $t \geq T = T(\|\varphi\|_\infty)$, where A is a constant independent of x and T .

Proof. This follows immediately from the above taking $\varepsilon = \|\varphi\|_\infty$. \square

Lemma 3.5 (Expectation of the population size). *Suppose $\beta \leq \lambda$. Then for all $x \in D$ and all $t \geq T = T(\|\varphi\|_\infty)$, we have*

$$\mathbb{E}_x[N_t] \leq A e^{(\beta - \lambda)t}$$

where A is a constant, independent of t and x .

Proof. This is a straightforward consequence of the Many-to-One Lemma and the above Corollary. \square

To prove almost sure extinction in the subcritical case, it is enough to show that $\mathbb{P}_x(N_t > 0) \rightarrow 0$ as $t \rightarrow \infty$. However, this is immediate from Lemma 3.5, since

$$\mathbb{P}_x(N_t > 0) \leq \mathbb{E}_x[N_t] \leq A e^{(\beta - \lambda)t}$$

which indeed tends to 0 in the case $\beta < \lambda$.

3.1.2 The Critical Case

When β is equal to λ , it is still the case that branching Brownian motion with parameter β dies out almost surely. However, since we can no longer rely on the fact that $e^{(\beta - \lambda)t} \rightarrow 0$ as $t \rightarrow \infty$, the decay of $\mathbb{E}_x[N_t]$ from Lemma 3.5 is lost, and we must apply a slightly more delicate argument. To improve the situation, we make use of the following Lemma, which can be found in [Wat65, Lemma 2.1]. We provide a proof for completeness.

Lemma 3.6 ([Wat65]). *For all $x \in D$*

$$\mathbb{P}_x(N_t \rightarrow 0 \text{ or } N_t \rightarrow \infty \text{ as } t \rightarrow \infty) = 1.$$

Proof. Since N_t is integer-valued it is sufficient to prove that $\mathbb{P}_x(N_t = k \text{ i.o.}) = 0$ for every $k \in \mathbb{N}$. Fix k and define a sequence of hitting and leaving times $(L_n, H_n)_{n \geq 1}$, by letting L_1 be the first time t that $N_t \neq N_0$, and H_1 be the first time that $N_t = k$. Then inductively, let L_n be the first time after H_{n-1} that $N_t \neq k$, and H_n the first time after this that $N_t = k$. We have to show that $\mathbb{P}_x(H_n < \infty) \rightarrow 0$ as $n \rightarrow \infty$. Set γ to be the infimum over all $x \in D$ of the probability that a Brownian motion started at x leaves the domain before an independent exponential waiting time. It is easily verified, using for example Corollary 3.4, that $\gamma > 0$. Then we have that $\mathbb{P}_x(H_1 < \infty) \leq \mathbb{P}(N_{L_1} \neq 0) \leq (1 - \gamma)$ and inductively, using the Markov property for the k particles alive at each time H_j , that $\mathbb{P}_x(H_n < \infty) \leq (1 - \gamma)(1 - \gamma^k)^{n-1}$. This completes the proof. \square

This, together with Lemma 3.3, provides the basis for the proof of Theorem 1.3 in the critical case. By Lemma 3.6, it will be sufficient to prove that $\mathbb{P}_x(N_t \rightarrow \infty \text{ as } t \rightarrow \infty) = 0$ for all $x \in D$. Since we know that $\sum_1^{N_t} \varphi(X_t^i) \rightarrow M_\infty < \infty$ almost surely, letting A_k be the event that $\{N_t \rightarrow \infty\} \cap \{\sum_1^{N_t} \varphi(X_t^i) \leq k \text{ eventually}\}$, we can write this probability as

$$\mathbb{P}_x(N_t \rightarrow \infty) = \sup_k \mathbb{P}_x(A_k)$$

since the A_k 's are increasing. Thus, it will be enough to show that $\mathbb{P}_x(A_k) = 0$ for all $k > 0$. To do this, fix k , and observe that $\mathbb{P}_x(A_k) = \lim_{m \rightarrow \infty} \mathbb{P}_x(A_k^m)$ where A_k^m is the event that $\{\{N_t \geq m\} \cap \{\sum_1^{N_t} \varphi(X_t^i) \leq k\} \text{ eventually}\}$. However, for

$$\{N_t \geq m\} \cap \left\{ \sum_1^{N_t} \varphi(X_t^i) \leq k \right\}$$

to occur, it must be the case that one of the particles in the system at time t has $\varphi(X_t^i) \leq \frac{k}{m}$ (since φ is positive.) Hence,

$$\mathbb{P}_x \left(\{N_t \geq m\} \cap \left\{ \sum_1^{N_t} \varphi(X_t^i) \leq k \right\} \right) \leq \mathbb{P}_x \left(\left(\sum_{i=1}^{N_t} \mathbb{1}_{\{\varphi(X_t^i) \leq \frac{k}{m}\}} \right) \geq 1 \right) \leq \mathbb{E}_x \left[\sum_{i=1}^{N_t} \mathbb{1}_{\{\varphi(X_t^i) \leq \frac{k}{m}\}} \right],$$

which is less than $K \frac{k}{m}$ for $t \geq T(k/m)$, as a consequence of Lemma 3.3. Since the probability of this holding for *all* large times is certainly smaller than the probability of it holding at time $T(k/m)$ say, we see that

$$\mathbb{P}_x(A_k) = \lim_{m \rightarrow \infty} \mathbb{P}_x(A_k^m) \leq \lim_{m \rightarrow \infty} K \frac{k}{m} = 0$$

as required.

Remark 3.7. *The proofs given above for the subcritical and critical cases rely purely on spectral properties of the Laplacian. These still hold for our more general diffusions, so no adaptation of the arguments is required.*

To complete the proof of Theorem 1.3 we must show that the decay of the survival probability in the critical case (the subcritical case having been dealt with by Lemma 3.5) is uniform in D . However, this will follow from elementary analysis, once we have noted that the survival probability is a solution of the FKPP equation in D . This was first observed by McKean, [McK75], in one dimension, but the proof extends directly to our situation. In fact, the relationship with the FKPP equation is the key tool employed in [Wat65] and [Sev58] to prove Theorem 1.3.

Lemma 3.8. *Assume that D satisfies Condition 2.1 and let $u(t, x) := \mathbb{P}_x(N_t > 0)$ for the critical system. Then $u \in C^{2,1}(D \times (0, \infty)) \cap C(\overline{D})$ is a solution of*

$$\begin{aligned} \partial u / \partial t &= \frac{1}{2} \Delta u + \lambda(u - u^2) && \text{on } D \times (0, \infty) \\ u(x, 0) &= \mathbb{1}_{\{x \in D\}} && \text{on } \overline{D} \times \{0\} \\ u(x, t) &= 0 && \text{on } \partial D \times (0, \infty). \end{aligned} \tag{3.2}$$

Proof. Conditioning on the first branching time of the process, we can write

$$u(x, t) = e^{-\lambda t} \mathbf{P}_x(\tau_D > t) + \int_0^t \lambda e^{-\lambda s} \mathbf{E}_x[(2u - u^2)(B_s, t - s) \mathbb{1}_{\{\tau_D > s\}}] ds.$$

An easy differentiation after making the change of variables $s \leftrightarrow t - s$ in the integral then provides (3.2). Note that $u(t, x) \rightarrow 0$ as $x \rightarrow \partial D$ since

$$\begin{aligned} \mathbb{P}_x(N_t > 0) &\leq 1 - \mathbb{P}_x(\text{the process becomes extinct before the first branching time}) \\ &\leq 1 - e^{-\lambda s_1} \mathbb{P}_x(\tau_D \leq s_1) \end{aligned}$$

for any s_1 , where the last line can be made arbitrarily small by first taking s_1 to 0, and then using the property (2) of Condition 2.1. Thus $u \in C(\overline{D})$. \square

To conclude the proof of 1.3, we note the continuous functions $u(x, t)$ are clearly decreasing in t and converge to the continuous function 0 for each $x \in \overline{D}$. This is a compact set and so by Dini's theorem, an elementary result from real analysis (see for example [Rud76, Theorem 7.13]), the decay must indeed be uniform.

Remark 3.9. *In the case of a more general diffusion with generator L and branching mechanism determined by offspring distribution A , the partial differential equation (3.2) becomes*

$$\frac{\partial u}{\partial t} = -Lu + \frac{\lambda}{m-1} ((1-u) - G(1-u))$$

where G is the probability generating function of A . This results in the same regularity for u .

4 Survival at Criticality: Proof of Theorem 1.4

Throughout this section, we will work in the critical case $\beta = \lambda$ (for binary branching Brownian motion) and also from now on assume the domain D to have C^1 boundary. We will prove the asymptotic (1.1) for the survival probability, using a combination of spine techniques, and analysis of the FKPP equation.

4.1 Spine Decomposition

It turns out that a helpful approach in the proof of (1.1) will be to change measure via the \mathbb{P}_x -martingale M_t from the previous section, see Lemma 2.3, and use a spine construction to describe the behaviour of λ -branching Brownian motion under this new measure. This is a so called *spine change of measure*, in that it changes the law of the initial particle, but then all subprocesses branching off this *spine* still behave as ordinary λ -branching Brownian motions.

To make sense of this, we extend our probability measure \mathbb{P}_x to a probability measure $\overline{\mathbb{P}}_x$ on a bigger space, by choosing one distinguished line of descent which we call the *spine*. We let the initial particle be part of the spine and then, whenever a spine particle splits, the new spine particle is chosen uniformly from its children. We denote the natural filtration of this new process by $(\overline{\mathcal{F}}_t)_{t \geq 0}$, and the position of the spine by $(\xi_t)_{t \geq 0}$. Then it follows from the fact that $e^{\lambda t} \varphi(B_t)/\varphi(x)$ is a martingale for our individual particle motion under \mathbf{P}_x (a consequence of (2.4)) that the process

$$\overline{M}_t = \frac{\varphi(\xi_t)}{\varphi(x)} 2^{S_t}$$

is a martingale with respect to $\overline{\mathcal{F}}_t$, where S_t is the number of branch points along the spine before time t . For a proof of this, see [Rob10, Theorem 2.4], and see also [Rob10] and [HH09] for details of the above construction. We can therefore define a new measure $\overline{\mathbb{Q}}_x$ on the same probability space as $\overline{\mathbb{P}}_x$ via

$$\left. \frac{d\overline{\mathbb{Q}}_x}{d\overline{\mathbb{P}}_x} \right|_{\overline{\mathcal{F}}_t} = \overline{M}_t.$$

Then there is a nice characterisation of the process under $\overline{\mathbb{Q}}_x$, which follows from the classical spine theory for such changes of measure developed in [HH09]; but see also [HH07], [CR88] and [HHK12] among others. Using the fact that under the change of measure for Brownian motion defined by

$$\left. \frac{d\mathbb{Q}_x}{d\mathbb{P}_x} \right|_{\mathcal{F}_t} = \frac{e^{\lambda t} \varphi(B_t)}{\varphi(x)} \quad (4.1)$$

we have that the new particle evolves as a Brownian motion conditioned to remain in D , see [Doo57], [Pin85], we obtain a description of the whole process under $\overline{\mathbb{Q}}_x$, as summarised in the Lemma below. We refer the reader to the papers cited above for more details, and proof of the characterisation.

Lemma 4.1. *Under the measure $\overline{\mathbb{Q}}_x$, the law of λ -branching Brownian motion with a distinguished spine can be constructed as follows:*

- *The initial particle evolves as a Brownian motion started at x and conditioned to remain in D for all time.*
- *At an accelerated rate of 2λ it splits into two particles.*
- *One of these particles, the spine particle, is chosen uniformly at random and goes on to repeat stochastically the behaviour of the initial ancestor.*
- *The other particle goes on to perform an independent λ -branching Brownian motion, starting from the point of fission.*

Alternatively, we can think of the process as being formed by a single spine particle, which evolves as a Brownian motion conditioned to remain in D , and along which ordinary λ -branching Brownian motions immigrate (branch off) at rate 2λ .

We also let $\mathbb{Q}_x := \overline{\mathbb{Q}}_x|_{\mathcal{F}_t}$ be the corresponding measure on the original filtration \mathcal{F}_t of the branching process. Then we have that

$$\left. \frac{d\mathbb{Q}_x}{d\mathbb{P}_x} \right|_{\mathcal{F}_t} = \frac{\sum_{i=1}^{N_t} \varphi(X_t^i)}{\varphi(x)} = \frac{M_t}{\varphi(x)} \quad (4.2)$$

(again see [Rob10, Theorem 2.4]). Furthermore, the probability under $\overline{\mathbb{Q}}_x$ that a certain particle X_t^j at time t is the spine particle, given \mathcal{F}_t , is equal to

$$\varphi(X_t^j) / \sum_{i=1}^{N_t} \varphi(X_t^i), \quad (4.3)$$

see [Rob10] or [HR14, Remark 1.2].

Remark 4.2. *For more general diffusions and branching mechanisms, we obtain a similar characterisation of the system after changing measure by the corresponding martingale. In this case, the spine particle will move under the law of the original diffusion conditioned to remain in the domain (defined by the same change of measure as in (4.1)), and at rate $\frac{m}{m-1}\lambda$, a certain number of the original branching diffusions will immigrate. The number of these immigrants has the size-biased offspring distribution \tilde{A} , where*

$$\mathbb{P}(\tilde{A} = k) = \frac{k+1}{m} \mathbb{P}(A = k+1)$$

for $k \geq 0$. Note the spine particle itself is not included in \tilde{A} .

One important fact we will use is that the position of the spine particle in the decomposition described by Lemma 4.1 converges quickly to an equilibrium distribution, with density φ^2 . In fact, if we assume that $\frac{1}{2}\Delta$ is *intrinsically ultracontractive* for the domain (for which a Lipschitz assumption is enough), it is well known that this convergence is uniform in the starting position.

Lemma 4.3. *Suppose that D is a bounded Lipschitz domain. If*

$$K^D(t, x, y) = \frac{e^{\lambda t} p^D(t, x, y) \varphi(y)}{\varphi(x)}$$

is the transition density for Brownian motion conditioned to remain in D , then for any $\varepsilon > 0$ there exists a constant C_ε depending only on the domain such that

$$\left| \frac{K^D(t, x, y)}{\varphi(y)^2} - 1 \right| \leq C_\varepsilon e^{-\gamma t}$$

for all $t > \varepsilon$ and $x, y \in D$ where $\gamma := \lambda_2 - \lambda_1 > 0$ is the spectral gap for $-1/2$ the Laplacian on D .

Proof. See for example [DS84] or [Bañ99, Equation (1.8)]. □

Remark 4.4. *The corresponding convergence to equilibrium when the single particle motion is governed by a generator L as in Theorem 1.3 still holds whenever the domain is Lipschitz (see [Bañ99].)*

4.2 Asymptotics for the Survival Probability

Using this spine decomposition, and the fact that the law of the spine particle converges to an equilibrium distribution as $t \rightarrow \infty$, we may first deduce that we have an asymptotic for the survival probability which is of the correct form.

Proposition 4.5. *For all $x \in D$*

$$\mathbb{P}_x(N_t > 0) \sim a(t) \varphi(x)$$

as $t \rightarrow \infty$, where

$$a(t) := \int_D \mathbb{P}_z(N_t > 0) \varphi(z) dz$$

converges to 0 as $t \rightarrow \infty$, and is independent of x .

Proof. The key idea for the proof of this is to write, recalling (4.2),

$$\frac{\mathbb{P}_x(N_t > 0)}{\varphi(x)} = \mathbb{Q}_x \left[\frac{1}{\sum_{i=1}^{N_t} \varphi(X_t^i)} \right]$$

and then show that the right hand side essentially does not depend on x for large t . The intuition behind this is that under the new measure, the position of the spine particle will converge very quickly to equilibrium. Then, contributions to the sum in the denominator from subprocesses branching off the spine before its position has become well mixed are unlikely to occur, as these have the law of standard λ -branching Brownian motions, which we know are unlikely to survive for a long time.

To begin, for $t_0 \leq t$, write

$$\sum_{i=1}^{N_t} \varphi(X_t^i) = M_t := M_{t_0, t} + M_{0, t_0}$$

where M_{0, t_0} is the sum of all contributions to M_t from subprocesses branching off the spine before time t_0 . Also define $f(r) := \mathbb{Q}_{\varphi^2}[1/\sum_{i=1}^{N_r} \varphi(X_r^i)]$, where φ^2 in the subscript indicates that the initial position is distributed according to the probability measure with density φ^2 . Note that this is a function of r only. Then for any $t_0 \leq t$

$$\frac{1}{M_t} = \frac{1}{M_{t_0, t}} + \left(\frac{1}{M_t} - \frac{1}{M_{t_0, t}} \right) = \frac{1}{M_{t_0, t}} - \frac{M_{0, t_0}}{M_t M_{t_0, t}}$$

and so

$$\mathbb{Q}_x \left[\frac{1}{\sum_{i=1}^{N_t} \varphi(X_t^i)} \right] = f(t - t_0) + (\mathbb{Q}_x \left[\frac{1}{M_{t_0,t}} \right] - f(t - t_0)) - \mathbb{Q}_x \left[\frac{M_{0,t_0}}{M_t M_{t_0,t}} \right]$$

where we label the error (second and third) terms above by $\epsilon_x^1(t, t_0)$ and $\epsilon_x^2(t, t_0)$ respectively. The plan is to show that we can choose $t_0(t) < t$ such that both the error terms become small as $t \rightarrow \infty$. The reason we expect these terms to decay is as in the heuristic discussion above, since the equilibrium distribution for Brownian motion conditioned to remain in D is precisely φ^2 .

Now observe that since the quantities whose expectations we are evaluating are \mathcal{F}_t measurable, their \mathbb{Q}_x and $\overline{\mathbb{Q}}_x$ expectations are the same, and we may work with either. Considering the definition of $M_{t_0,t}$, and conditioning on \mathcal{G}_{t_0} , for $(\mathcal{G}_s)_{s \geq 0}$ the filtration generated by the position of the spine up to time s (a subfiltration of $(\overline{\mathcal{F}}_s)_{s \geq 0}$) we see that

$$\epsilon_x^1(t, t_0) = \overline{\mathbb{Q}}_x[\overline{\mathbb{Q}}_x[1/M_{t_0,t}|\mathcal{G}_{t_0}]] - f(t - t_0)$$

is simply the difference in expectation of the function of y ,

$$\mathbb{Q}_y \left[\frac{1}{\sum_{i=1}^{N_{t-t_0}} \varphi(X_{t-t_0}^i)} \right] = \mathbb{P}_y(N_{t-t_0} > 0) / \varphi(y)$$

under the $\overline{\mathbb{Q}}_x$ law of the spine particle at time t_0 , and the law with density φ^2 . As t_0 increases, we know that the law of the spine particle approaches the law with this density, and by Lemma 4.3 we can quantify this with the bound

$$\sup_{x,y \in D} \left| \frac{K^D(t_0, x, y)}{\varphi(y)^2} - 1 \right| \leq C e^{-\gamma t_0} \quad (4.4)$$

for all $t_0 > 1$ say, where $K^D(s, x, y)$ is the transition density of Brownian motion conditioned to remain in D . Hence we have the estimate, for all $t_0 > 1$:

$$|\epsilon_x^1(t, t_0)| \leq \sup_{x,y \in D} \left| \frac{K^D(t_0, x, y)}{\varphi(y)^2} - 1 \right| \int_D \varphi(y) \mathbb{P}_y(N_{t-t_0} > 0) dy \leq C e^{-\gamma t_0} f(t - t_0). \quad (4.5)$$

To bound the second term, write $A_{t_0}^t$ for the event that some subprocess branching from the spine before time t_0 survives until (total) time t . Since $M_{0,t_0}/M_t M_{t_0,t}$ is positive and less than or equal to $1/M_{t_0,t}$ we see that

$$|\epsilon_x^2(t, t_0)| \leq \overline{\mathbb{Q}}_x \left[\frac{1}{M_{t_0,t}} \mathbb{1}_{A_{t_0}^t} \right].$$

Again, to estimate this we condition; but now on $\tilde{\mathcal{G}}_{t_0}$, where $\tilde{\mathcal{G}} \supset \mathcal{G}$ is the filtration which also contains information about the branching points along the spine. The reason for doing this is that we know, given the position of the spine (ξ_s) for $0 \leq s \leq t_0$ and all its branching points, that the subprocesses branching off the spine before t_0 and the process continuing on from ξ_{t_0} are independent. Thus, the term on the left above is equal to

$$\overline{\mathbb{Q}}_x \left[\overline{\mathbb{Q}}_x \left[\mathbb{1}_{A_{t_0}^t} | \tilde{\mathcal{G}}_{t_0} \right] \overline{\mathbb{Q}}_x \left[\frac{1}{M_{t_0,t}} | \tilde{\mathcal{G}}_{t_0} \right] \right] = \overline{\mathbb{Q}}_x \left[\overline{\mathbb{Q}}_x \left[\mathbb{1}_{A_{t_0}^t} | \tilde{\mathcal{G}}_{t_0} \right] \frac{\mathbb{P}_{\xi_{t_0}}(N_{t-t_0} > 0)}{\varphi(\xi_{t_0})} \right],$$

where we can now show that the conditional probability $\overline{\mathbb{Q}}_x[\mathbb{1}_{A_{t_0}^t} | \tilde{\mathcal{G}}_{t_0}]$ is small. Indeed, since the probability that any subprocess branching off the spine before t_0 survives until total time t is less than $\sup_{w \in D} |\mathbb{P}_w(N_{t-t_0} > 0)|$, we see that

$$\overline{\mathbb{Q}}_x \left[\mathbb{1}_{A_{t_0}^t} | \tilde{\mathcal{G}}_{t_0} \right] \leq S_{t_0} \sup_{w \in D} |\mathbb{P}_w(N_{t-t_0} > 0)|,$$

where S_{t_0} is the number of such subprocesses. Moreover, again for $t_0 > 1$ say by (4.4), we have

$$\overline{\mathbb{Q}}_x \left[\frac{\mathbb{P}_{\xi_{t_0}}(N_{t-t_0} > 0)}{\varphi(\xi_{t_0})} \right] \lesssim f(t - t_0)$$

where the implied constant depends only on D . Combining all of the above, and noting that S_{t_0} is independent of the motion under $\overline{\mathbb{Q}}_x$ with $\overline{\mathbb{Q}}_x[S_{t_0}] = 2\lambda t_0$ provides the final estimate

$$|\epsilon_x^2(t, t_0)| \leq \tilde{C} t_0 f(t - t_0) \sup_{w \in D} |\mathbb{P}_w(N_{t-t_0} > 0)| \quad (4.6)$$

for all $t_0 > 1$, where \tilde{C} is another constant. With both these error bounds in hand, we may deduce that

$$\left| \frac{\mathbb{P}_x(N_t > 0) / \varphi(x)}{f(t - t_0)} - 1 \right| \leq \left| \frac{\epsilon_x^1(t, t_0)}{f(t - t_0)} \right| + \left| \frac{\epsilon_x^2(t, t_0)}{f(t - t_0)} \right| \leq C e^{-\gamma t_0} + \tilde{C} t_0 \sup_{w \in D} |\mathbb{P}_w(N_{t-t_0} > 0)|$$

for any $x \in D$, and $1 < t_0 < t$.

Now, since we know that $\sup_{w \in D} |\mathbb{P}_w(N_{t-t_0} > 0)| \rightarrow 0$ as $s \rightarrow \infty$, it is possible to choose $t_0(t)$ such that both $t_0(t) \rightarrow \infty$ and $t_0(t) \sup_{w \in D} |\mathbb{P}_w(N_{t-t_0} > 0)| \rightarrow 0$ as $t \rightarrow \infty$. Then we have, letting $c(t) = f(t - t_0(t))$, that

$$\left| \frac{\mathbb{P}_x(N_t > 0) / \varphi(x)}{c(t)} - 1 \right| \rightarrow 0 \quad (4.7)$$

as $t \rightarrow \infty$, uniformly for $x \in D$. To complete the proof, we need only show that $c(t)$ must be asymptotically equivalent to $a(t) := \int_D \mathbb{P}_x(N_t > 0) \varphi(x) dx$ as $t \rightarrow \infty$. Note that $a(t)$ is less than $\sup_{w \in D} |\mathbb{P}_w(N_{t-t_0} > 0)|$, and so clearly decays with t . To see the equivalence, observe that

$$\left| \frac{a(t)}{c(t)} - 1 \right| = \left| \int_D \left(\frac{\mathbb{P}_x(N_t > 0)}{c(t)} - \varphi(x) \right) \varphi(x) dx \right| \leq \left\| \frac{\mathbb{P}_x(N_t > 0)}{c(t)} - \varphi(x) \right\|_{L^2(D)} \quad (4.8)$$

where the inequality holds by Cauchy-Schwarz and the fact that $\int_D \varphi^2 = 1$. Multiplying (4.7) by $\varphi(x)$ and integrating tells us that the final expression converges to 0 as $t \rightarrow \infty$. \square

Remark 4.6. *The proof of this Lemma remains essentially the same in our more general framework, using Remark 4.2 in place of Lemma 4.3. Some more care is required to bound*

$$\overline{\mathbb{Q}}_x[S_{t_0}],$$

as multiple processes may now immigrate at each branching point on the spine, but since the size-biased distribution has finite mean (we are assuming that A has finite variance) this is again less than some constant times t_0 .

4.3 Asymptotics for $a(t)$

Now, since by Lemma 3.8 we know that

$$\mathbb{P}_x(N_t > 0) := u(x, t)$$

is a solution of the FKPP equation in D , we can find an ODE which is satisfied by $a(t)$. This will allow us to deduce the desired asymptotic for $a(t)$ as $t \rightarrow \infty$. More precisely, we have the following:

Lemma 4.7. *Assume that D is C^1 . Letting $a(t) = \int_D u(x, t) \varphi(x) dx$ we have that $a(t)$ is differentiable for all $t > 0$ and*

$$\frac{da}{dt}(t) = -\lambda \int_D u^2(x, t) \varphi(x) dx. \quad (4.9)$$

Proof. First suppose that $u(\cdot, t) \in H_0^1(D)$ and $\Delta u(\cdot, t) \in L^2(D)$ for all $t > 0$. Then, since

$$\frac{\partial u}{\partial t}(x, t) = \frac{1}{2} \Delta u(x, t) + \lambda(u(x, t) - u^2(x, t))$$

we see that $\frac{\partial u}{\partial t}(\cdot, s) \in L^2(D)$ for all strictly positive s , and

$$\int_D \frac{\partial u}{\partial t}(x, s) \varphi(x) dx = \int_D \left(\frac{1}{2} \Delta u(x, s) + \lambda(u(x, s) - u^2(x, s)) \right) \varphi(x) dx$$

is well defined. Furthermore, we have that $\varphi \in H_0^1(D)$ (since ∂D is assumed C^1 and φ vanishes on the boundary - see [Eva98, §5.5]) and that $u \in H_0^1(D)$ by assumption. This means that we can integrate by parts against $\varphi \in H_0^1(D)$ and use that φ is an eigenfunction of the Laplacian to obtain the equality

$$\int_D \frac{du}{dt}(x, s) \varphi(x) ds = -\lambda \int_D u^2(x, s) \varphi(x) dx. \quad (4.10)$$

Observe that the left hand side is continuous in s (for $s > 0$) since the right hand side must be by continuity of u^2 and dominated convergence.

Hence, for any $t > 0$, letting $0 < a < t$ we see that

$$-\lambda \int_D u^2(x, t) \varphi(x) dx = \frac{d}{dt} \int_a^t \int_D -\lambda u^2(x, s) \varphi(x) dx = \frac{d}{dt} \int_a^t \int_D \frac{\partial u}{\partial t}(x, s) \varphi(x) dx ds$$

by the continuity discussed above, and (4.10). Then, applying Fubini and using the continuity of $\frac{du}{dt}(x, s)$ in s for fixed x , we can write this as

$$\frac{d}{dt} \int_D \int_a^t \frac{\partial u}{\partial t}(x, s) ds \varphi(x) dx = \frac{d}{dt} \int_D u(x, t) \varphi(x) - u(x, a) \varphi(x) dx = \frac{da}{dt}(t).$$

Therefore, we need only show that $u \in H_0^1(D)$ and $\Delta u \in L^2(D)$. However, this is simply the regularity that u obtains by virtue of being a solution of (3.2). This is a straightforward consequence of the standard regularity theory for parabolic PDEs. Since the domain is C^1 and u vanishes on the boundary, it is enough to show that

$$\sum_{k=1}^{\infty} \lambda_k(f, \varphi_k)^2 < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \lambda_k^2(f, \varphi_k)^2 < \infty \quad (4.11)$$

(see, for example, [Tho06, Lemma 3.1]). This can be proved using the Duhamel representation for u as a solution of (3.2) and a standard bootstrapping argument. We omit the straightforward calculations. \square

Remark 4.8. *For the more general branching diffusion we can apply the same arguments to show that*

$$\frac{da}{dt}(t) = -\frac{\lambda}{m-1} \int_D (G(1-u) + mu - 1) \varphi(x) dx \quad (4.12)$$

where G is the probability generating function of A . Continuity of the right-hand side requires an extra application of the dominated convergence theorem to first see that $G(1-u) = \mathbb{E}[(1-u)^A]$ is continuous. To bound the Sobolev norms in (4.11) one must observe that

$$\nabla(G(1-u)) = -\nabla u \mathbb{E}[A(1-u)^A]$$

is bounded in modulus by a constant times $|\nabla u|$.

This allows us to deduce an asymptotic for $a(t)$, which completes the proof of Theorem 1.4.

Proposition 4.9.

$$a(t) \underset{t \rightarrow \infty}{\sim} \frac{1}{(\lambda \int_D \varphi(y)^3 dy) t}.$$

Proof. The desired asymptotic for $a(t)$ follows fairly easily from Lemma 4.7 since we have, writing $u(t, x) = a(t)\varphi(x) + v(t, x)$ and substituting this into (4.9), that

$$\frac{da}{dt}(t) = -\lambda a^2(t) \int_D \varphi(y)^3 dy - 2\lambda a(t) \int_D v(t, x) \varphi^2(x) dx - \lambda \int_D v^2(t, x) \varphi(x) dx. \quad (4.13)$$

Then, Lemma 4.5 tells us that the second two terms are negligible compared with the first for large t . Indeed, since $|u(t, x)/a(t)\varphi(x) - 1| = |v(t, x)/a(t)\varphi(x)| \rightarrow 0$ uniformly in x , we have that

$$\frac{da}{dt}(t) = \left(-\lambda \int_D \varphi(y)^3 dy - g(t) \right) a^2(t) \quad \text{for} \quad g(t) := 2\lambda \int_D \frac{v(t, x)}{a(t)} \varphi^2(x) dx + \lambda \int_D \frac{v(t, x)^2}{a(t)^2} \varphi(x) dx$$

where $g(t) \rightarrow 0$ as $t \rightarrow \infty$. Thus we obtain, denoting differentiation with respect to t by a dot, that

$$-\lambda \int_D \varphi(y)^3 dy - |g(t)| \leq \frac{\dot{a}(t)}{a^2(t)} \leq -\lambda \int_D \varphi(y)^3 dy + |g(t)|.$$

Moreover, since $\frac{da^{-1}}{dt}(t) = -\frac{\dot{a}(t)}{a^2(t)}$, integration yields that

$$\left(\lambda \int_D \varphi(y)^3 dy \right) (t-1) - \int_1^t |g(s)| ds + \frac{1}{a(1)} \leq \frac{1}{a(t)} \leq \left(\lambda \int_D \varphi(y)^3 dy \right) (t-1) + \int_1^t |g(s)| ds + \frac{1}{a(1)}$$

where we have started from 1 to avoid any differentiability issues at 0. Note that $\dot{a}(s)/a^2(s)$ is clearly integrable over $[1, t]$ for any t , and so we are justified in applying the fundamental theorem of calculus here. Upon dividing by $(\lambda \int_D \varphi(y)^3 dy) t$ we see that

$$\left| \frac{\frac{1}{(\lambda \int_D \varphi(y)^3 dy)t}}{a(t)} - 1 \right| \leq \frac{1}{t} + \frac{1}{(\lambda \int_D \varphi(y)^3 dy)} \left(\frac{1}{a(1)t} + \frac{\int_0^t |g(s)| ds}{t} \right).$$

The first term in the brackets on the right hand side of this expression clearly converges to 0 as $t \rightarrow \infty$. Furthermore, since $|g|$ is bounded and $|g(s)| \rightarrow 0$ as $s \rightarrow \infty$, so does the second. This yields the result. \square

Remark 4.10. Note that this Proposition, combined with the proof Theorem 1.3, in particular (4.7), shows that

$$\left| \frac{t\lambda \int_D \varphi(y)^3 dy \times \mathbb{P}_x(N_t > 0)}{\varphi(x)} - 1 \right| \rightarrow 0$$

as $t \rightarrow \infty$, uniformly in x .

Remark 4.11. For the more general set up, observe that

$$G(1-u) + mu - 1 = \frac{\mathbb{E}[A^2] - \mathbb{E}[A]}{2} u^2 + o(u^2)$$

by Taylor's theorem and our moment assumption on the offspring distribution A (see [DN80, Theorem A] for a general statement concerning Taylor expansions of probability generating functions). Replacing the expression for $\frac{da}{dt}$ by (4.12) and proceeding as above, we may incorporate the error term from the integral into $g(t)$, and reach the desired conclusion.

5 The Conditioned System

Theorem 1.4 allows us to study the law of branching Brownian motion conditioned to survive for a long time in much greater depth. One aspect of the limiting behaviour is captured by what happens to the law of the process run up to some fixed time T , if it is then conditioned to survive until a much larger time t . It turns out that this limiting description is given precisely by the evolution of the process under \mathbb{Q}_x , as described in Lemma 4.1.

Proof of Proposition 1.5. Recall, we would like to prove that for any $T \geq 0$, $x \in D$ and $B \in \mathcal{F}_T$, we have that

$$\lim_{t \rightarrow \infty} \mathbb{P}_x(B|N_t > 0) = \mathbb{Q}_x(B).$$

Conditioning on \mathcal{F}_T , we see that

$$\mathbb{P}_x(B|N_t > 0) = \frac{\mathbb{E}_x[\mathbb{1}_B \mathbb{P}_x(N_t > 0|\mathcal{F}_T)]}{\mathbb{P}_x(N_t > 0)} := \frac{\mathbb{E}_x[\mathbb{1}_B Y]}{\mathbb{P}_x(N_t > 0)}$$

where we have defined

$$Y := \mathbb{P}_x(N_t > 0|\mathcal{F}_T) = \sum_{i=1}^{N_T} \mathbb{P}_{X_T^i}(N_{t-T} > 0) \left(\prod_{j < i} \mathbb{P}_{X_T^j}(N_{t-T} = 0) \right).$$

Then, from our asymptotic for the survival probability and the fact that $\frac{t}{t-T} \rightarrow 1$ as $t \rightarrow \infty$, it follows that

$$\frac{\mathbb{1}_B Y}{\mathbb{P}_x(N_t > 0)} \xrightarrow{t \rightarrow \infty} \frac{\sum_{i=1}^{N_T} \varphi(X_T^i)}{\varphi(x)} \mathbb{1}_B = \frac{M_T}{M_0} \mathbb{1}_B$$

almost surely, as $t \rightarrow \infty$. Moreover, we have that $Y \leq \sum_{i=1}^{N_T} \mathbb{P}_{X_T^i}(N_{t-T} > 0) \lesssim \frac{M_T}{t-T}$ for all large enough t (recalling that our asymptotic estimates were uniform in D), and so we can dominate $\mathbb{1}_B Y / \mathbb{P}_x(N_t > 0)$ by an integrable random variable, namely a constant multiple of M_T . The dominated convergence theorem then provides the result. \square

Given the asymptotic for the survival probability, it is also not too much work to prove Theorem 1.6, which gives some limiting information on the positions of particles at time t , given survival. Recall (since we are working in the branching Brownian motion case) we would like to show that for any f with $(f^2, \varphi) < \infty$ that we have

$$(t^{-1} \sum_{i=1}^{N_t} f(X_t^i) | N_t > 0) \rightarrow Z$$

in distribution as $t \rightarrow \infty$, where $Z \sim \text{Exp}(1/\lambda(\varphi, f) \int_D \varphi(y)^3 dy)$. To prove Theorem 1.6, we will use the method of moments, relying on Theorem 1.4 and the following Lemma.

Lemma 5.1. *For all $n \in \mathbb{N}$ and $x \in D$,*

$$\frac{\mathbb{E}_x \left[\left(\sum_{i=1}^{N_t} \varphi(X_t^i) \right)^n \right]}{n! \varphi(x) \lambda^{n-1} \left(\int_D \varphi^3 \right)^{n-1} t^{n-1}} \rightarrow 1$$

as $t \rightarrow \infty$, uniformly in x .

Proof of Lemma 5.1. The proof of this relies on the expression for $\mathbb{E}_x[(\sum_{i=1}^{N_t} \varphi(X_t^i))^n]$ that one obtains from a Many-To-Few generalisation of the Many-to-Two Lemma, see below, and then proceeds by induction. We start with $n = 1$. This case is simple, however, because we know that

$$\left| \mathbb{E}_x \left[\frac{\sum_{i=1}^{N_t} \varphi(X_t^i)}{\varphi(x)} \right] - 1 \right| = 0$$

for all t and x , since the expectation is that of our mean 1 martingale M_t/M_0 .

We also record here that

$$\left| \frac{e^{\lambda t} \mathbf{E}_x [\varphi^2(B_t) \mathbb{1}_{\{t > \tau_D\}}]}{\varphi(x) \int_D \varphi^3} - 1 \right| = \frac{1}{\int_D \varphi(y)^3 dy} \int_D \frac{e^{\lambda t} p^D(t, x, y) \varphi^2(y)}{\varphi(x)} - \varphi(y)^3 dy \leq \sup_{y \in D} \left| \frac{K^D(t, x, y)}{\varphi(y)^2} - 1 \right| \quad (5.1)$$

which we know is less than $C e^{-\gamma t}$ for all x and $t \geq 1$ say, where C is some universal constant. This fact will be crucial to the induction.

For the inductive step, we need a Many-to-Few Lemma, which is a generalisation of Lemma 3.1. This tells us, as a special case, that

$$\mathbf{E}_x \left[\left(\sum_{i=1}^{N_t} \varphi(X_t^i) \right)^n \right] = e^{\lambda t} \mathbf{E}_x [\varphi(B_t)^n \mathbb{1}_{\{t > \tau_D\}}] + \sum_{j=1}^{n-1} \binom{n}{j} \int_0^t \lambda e^{\lambda s} \mathbf{E}_x [\mathbb{1}_{\{t > s\}} E^j(t-s, B_s)] ds \quad (5.2)$$

where $E^j(s, x) := \mathbf{E}_x[(\sum_{i=1}^{N_s} \varphi(X_s^i))^{n-j}] \mathbf{E}_x[(\sum_{i=1}^{N_s} \varphi(X_s^i))^j]$, and is proved in a very general setting in [HR15, Lemma 1]. Thus, for $n > 1$ we can break up

$$\left| \frac{\mathbf{E}_x \left[\left(\sum_{i=1}^{N_t} \varphi(X_t^i) \right)^n \right]}{n! \varphi(x) \lambda^{n-1} (\int_D \varphi^3)^{n-1} t^{n-1}} - 1 \right|$$

into n parts, each corresponding to one of the terms in (5.2). Since we know that $e^{\lambda t} \mathbf{E}_x [\varphi(B_t)^n \mathbb{1}_{\{t > \tau_D\}}] \leq \|\varphi\|_\infty^{n-1} \varphi(x)$, the first of these terms will tend to 0 uniformly in x as $t \rightarrow \infty$. Thus we need only show that for each $1 \leq j \leq n-1$, we have

$$\left| \frac{\int_0^t e^{\lambda s} \mathbf{E}_x [\mathbb{1}_{\{t > s\}} E^j(t-s, B_s)] ds}{j!(n-j)! \varphi(x) \lambda^{n-2} (\int_D \varphi^3)^{n-1} t^{n-1}} - \frac{1}{n-1} \right| \rightarrow 0 \quad (5.3)$$

uniformly in x as $t \rightarrow \infty$. In the following we set

$$Q_y^k(r) := \frac{\mathbf{E}_y \left[\left(\sum_{i=1}^{N_r} \varphi(X_r^i) \right)^k \right]}{k! \lambda^{k-1} (\int_D \varphi^3)^{k-1} r^{k-1}} \quad (5.4)$$

for $y \in D$, $k \in \mathbb{N}$ and $r \geq 0$ so that (5.3) is less than

$$\left| \frac{\int_0^t e^{\lambda s} (t-s)^{n-2} \mathbf{E}_x \left[\mathbb{1}_{\{t > s\}} \left(Q_{B_s}^{n-j}(t-s) Q_{B_s}^j(t-s) - \varphi^2(B_s) \right) \right] ds}{t^{n-1} \varphi(x) \int_D \varphi^3} \right| + \left| \frac{\int_0^t e^{\lambda s} \mathbf{E}_x [\varphi^2(B_s) \mathbb{1}_{\{t > s\}}] (t-s)^{n-2} ds}{t^{n-1} \varphi(x) \int_D \varphi^3} - \frac{1}{n-1} \right|. \quad (5.5)$$

We will show that both the terms here converge to 0, uniformly in x .

The second expression is relatively easy to deal with, since $\int_0^t (t-s)^{n-2} / t^{n-1} = 1/(n-1)$. This means that we can pull the $1/(n-1)$ term, and then the modulus, inside of the integral. After doing this, and noting that $(t-s)^{n-2} / t^{n-1} \leq 1/t$ for all $s \in [0, t]$, we see that the term is bounded above by

$$\frac{1}{t} \int_0^t \left| \frac{e^{\lambda s} \mathbf{E}_x [\varphi^2(B_s) \mathbb{1}_{\{t > s\}}]}{\varphi(x) \int_D \varphi^3} - 1 \right| ds.$$

From here the result follows by (5.1). Note that the bound $\mathbf{E}_x[\varphi^2(B_s) \mathbb{1}_{\{t > s\}}] \leq e^{-\lambda s} \varphi(x) \|\varphi\|_\infty$ (which holds for all s and x) tells us that the integrand is uniformly bounded in x .

For the first expression, we use our induction hypothesis. This tells us that $Q_z^{n-j}(r)Q_z^j(r)/\varphi(z)^2 \rightarrow 1$ uniformly in z as $r \rightarrow \infty$. It is also clear from the definition (5.4) that for fixed n and j , $r^{n-2}Q_z^{n-j}(r)Q_z^j(r)$ is uniformly bounded on any compact interval $[0, T]$ (for example by bounding the expectation in (5.4) by the corresponding expectation for a Yule process and using that φ is bounded). With this in mind we bound (5.5) above by

$$\begin{aligned} & \frac{\sup_z \sup_{r \geq T} |Q_z^{n-j}(r)Q_z^j(r)/\varphi(z)^2 - 1|}{t \int_D \varphi(y)^3 dy} \int_0^{t-T} e^{\lambda s} \frac{\mathbf{E}_x [\mathbb{1}_{\{\tau_s > 0\}} \varphi^2(B_s)]}{\varphi(x)} ds \\ & + \frac{\sup_z \sup_{r \leq T} (|r^{n-2}Q_z^{n-j}(r)Q_z^j(r)| + |r^{n-2}\varphi(z)^2|)}{t^{n-1} \int_D \varphi(y)^3 dy} \int_{t-T}^t e^{\lambda s} \frac{\mathbf{E}_x [\mathbb{1}_{\{\tau_D > s\}}]}{\varphi(x)} ds \end{aligned}$$

where we have used that $(t-s)^{n-2}/t^{n-1} \leq 1/t$ for the first part. Since

$$e^{\lambda s} \frac{\mathbf{E}_x [\mathbb{1}_{\{\tau_D > s\}}]}{\varphi(x)} = \int_D \frac{e^{\lambda s} p^D(s, x, y)}{\varphi(x)\varphi(y)} \varphi(y) dy$$

is the expectation of $\varphi(\xi_s)$ for ξ a Brownian motion conditioned to remain in D , and therefore bounded uniformly in s and x , we obtain the convergence to 0 by letting first $T \rightarrow \infty$ and then $t \rightarrow \infty$ in the above. \square

Proof of Theorem 1.6. Lemma 5.1 combined with Theorem 1.4 tells us that for each $n > 0$,

$$\mathbb{E}_x[(t^{-1} \sum_{i=1}^{N_t} \varphi(X_t^i))^n | N_t > 0] \rightarrow n! \left(\lambda \int_D \varphi^3 \right)^n$$

as $t \rightarrow \infty$, where the right hand side is the n th moment of an $\text{Exp}(1/\lambda \int_D \varphi^3)$ random variable. Since convergence of the moments is enough to ensure convergence in distribution when the limiting distribution is nice enough, see for example [Bil95, Theorem 30.2], Theorem 1.6 is proved in the case $f = \varphi$. To deal with general f we write $\tilde{f} = f - (\varphi, f) \varphi$. We will show that for any $\varepsilon > 0$

$$\mathbb{P}_x(|t^{-1} \sum_{i=1}^{N_t} \tilde{f}(X_t^i)| > \varepsilon | N_t > 0) \rightarrow 0$$

as $t \rightarrow \infty$ (uniformly in x), which implies the result by the above decomposition and the proof when $f = \varphi$. To do this we use Markov's inequality and the Many-to-Two Lemma. This Lemma tells us that $\mathbb{E}_x[(t^{-1} \sum_{i=1}^{N_t} \tilde{f}(X_t^i))^2 | N_t > 0]$ is equal to

$$\frac{\varphi(x)}{t \mathbb{P}_x(N_t > 0)} \left(\frac{e^{\lambda t} \mathbf{E}_x[\tilde{f}(B_t)^2 \mathbb{1}_{\{\tau_D > t\}}] + 2 \int_0^t \lambda e^{\lambda s} \mathbf{E}_x[\mathbb{1}_{\{\tau_D > s\}} \mathbb{E}_{B_s}[\sum_{i=1}^{N_{t-s}} \tilde{f}(X_{t-s}^i)]^2] ds}{t \varphi(x)} \right)$$

where we know that the expression outside the brackets is uniformly bounded in t and x (for t large enough). We also know that

$$\frac{e^{\lambda t} \mathbf{E}_x[\tilde{f}(B_t)^2 \mathbb{1}_{\{\tau_D > t\}}]}{t \varphi(x)} \leq t^{-1} \left(1 + \sup_y \left| \frac{K^D(t, x, y)}{\varphi(y)^2} - 1 \right| \right) (\varphi, \tilde{f}^2) \quad (5.6)$$

and so this term converges uniformly to 0 by the assumption on f . To conclude we use the fact that $(\varphi, \tilde{f}) = 0$, which was the reason for choosing \tilde{f} as we did. This means that $\sup_z \varphi(z)^{-1} \mathbb{E}_z[\sum_{i=1}^{N_r} \tilde{f}(X_r^i)] \rightarrow 0$ as $r \rightarrow \infty$, by the Many-to-One Lemma and the same argument used for the bound in (5.6). Thus we have that

$$\begin{aligned} (t \varphi(x))^{-1} \int_0^t \lambda e^{\lambda s} \mathbf{E}_x[\mathbb{1}_{\{\tau_D > s\}} \mathbb{E}_{B_s}[\sum_{i=1}^{N_{t-s}} \tilde{f}(X_{t-s}^i)]^2] ds & \leq \frac{\int_{t-T}^t \lambda e^{\lambda s} \mathbf{E}_x[\mathbb{1}_{\{\tau_D > s\}} \mathbb{E}_{B_s}[\sum_{i=1}^{N_{t-s}} \tilde{f}(X_{t-s}^i)]^2] ds}{t \varphi(x)} \\ & + \sup_{r \geq T} \sup_z \left(\varphi(z)^{-1} \mathbb{E}_z[\sum_{i=1}^{N_r} \tilde{f}(X_r^i)] \right)^2 \frac{\int_0^{t-T} \lambda e^{\lambda s} \mathbf{E}_x[\varphi(B_s)^2 \mathbb{1}_{\{\tau_D > t\}}] ds}{t \varphi(x)} \end{aligned}$$

where the second term on the right can be made arbitrarily small for all $t \geq T$ as long as T is large enough (using the bound $e^{\lambda s} \mathbf{E}_x[\varphi(B_s)^2 \mathbb{1}_{\{\tau_D > t\}}] \leq \varphi(x) \|\varphi\|_\infty$ on the integrand). Therefore, we are left with having to show that, for fixed T , the first term converges to 0 (uniformly in x) as $t \rightarrow \infty$. However, we can apply the Many-to-one Lemma and Cauchy-Schwarz to the $\mathbf{E}_{B_s}[\cdot]$ term in the integral, and then the Markov property, to see that

$$\mathbf{E}_x[\mathbb{1}_{\{\tau_D > s\}} \mathbf{E}_{B_s}[\sum_{i=1}^{N_{t-s}} \tilde{f}(X_{t-s}^i)]^2] \leq \mathbf{E}_x[\mathbb{1}_{\{\tau_D > s\}} e^{2\lambda(t-s)} \mathbf{E}_{B_s}[\tilde{f}^2(\tilde{B}_{t-s})]] \leq e^{2\lambda(t-s)} \mathbf{E}_x[\mathbb{1}_{\{\tau_D > t\}} \tilde{f}(B_t)^2]$$

where the process \tilde{B} in the middle expression is just an independent Brownian motion. Using the assumption that $(\varphi, \tilde{f}^2) < \infty$ and (5.6) once more, the result follows from computing the integral. \square

Remark 5.2. We note here that Lemma 5.1 would also hold if we started the initial particle in a random position, with density φ^2 . This follows directly from the proof, since everything is uniform in x . In fact, since the asymptotic for the survival probability is also uniform in the starting point by Remark 4.10, we have that Theorem 1.6 holds when $f = \varphi$ and we start the system in this random initial position.

We conclude by explaining how one can obtain Corollary 1.7 from here, which describes the asymptotic distribution of a particle picked at random from the population, given survival.

Proof of Corollary 1.7. To prove the Corollary we first show that for any f with $(\varphi, f^2) < \infty$

$$\mathbb{P}_x \left(\left| \frac{\sum_{i=1}^{N_t} f(X_t^i)}{\sum_{i=1}^{N_t} \varphi(X_t^i)} - (\varphi, f) \right| > \varepsilon \mid N_t > 0 \right) \rightarrow 0 \quad (5.7)$$

as $t \rightarrow \infty$. Defining \tilde{f} as in the proof of Theorem 1.6, this is equal to

$$\mathbb{P}_x \left(\left| \frac{t \sum_{i=1}^{N_t} \tilde{f}(X_t^i)}{t \sum_{i=1}^{N_t} \varphi(X_t^i)} \right| > \varepsilon \mid N_t > 0 \right) \leq \mathbb{P}_x \left(t^{-1} \sum_{i=1}^{N_t} \tilde{f}(X_t^i) > \delta \mid N_t > 0 \right) + \mathbb{P}_x \left(t^{-1} \sum_{i=1}^{N_t} \varphi(X_t^i) < \delta/\varepsilon \mid N_t > 0 \right) \quad (5.8)$$

for any $\delta > 0$. From the proof of Theorem 1.6, if we take a limit as $t \rightarrow \infty$ on the right hand side, we are left with simply the probability that an exponential random variable is less than δ/ε . Taking $\delta \rightarrow 0$ proves (5.7). The Corollary then follows by applying the above with both f and the constant function 1, and writing $\sum f(X_t^i)/N_t = \sum f(X_t^i)/\sum \varphi(X_t^i) \times \sum \varphi(X_t^i)/N_t$. \square

Corollary 5.3. For f as in Corollary 1.7, and any $\varepsilon > 0$, we have that

$$\mathbb{P}_x \left(\left| \frac{\sum_{i=1}^{N_t} f(X_t^i)}{\sum_{i=1}^{N_t} \varphi(X_t^i)} - (\varphi, f) \right| > \varepsilon \mid N_t > 0 \right) \rightarrow 0 \quad (5.9)$$

as $t \rightarrow \infty$, uniformly in the starting position x .

Proof. Note that the proof of Theorem 1.6 immediately gives us convergence of the first term in (5.8) to 0 as $t \rightarrow \infty$, for any $\delta > 0$, uniformly in x . Therefore, it is sufficient for us to prove that there exists a function $g(\delta)$ converging to 0 as $\delta \downarrow 0$, such that for every $\eta > 0$ there exists a T with

$$\mathbb{P}_x \left(t^{-1} \sum_{i=1}^{N_t} \varphi(X_t^i) < \delta/\varepsilon \mid N_t > 0 \right) \leq g(\delta) + \eta$$

for all $t \geq T$ and all $x \in D$. To do this, we change measure to \mathbb{Q}_x , and use the fact that the spine particle under \mathbb{Q}_x converges uniformly to an equilibrium distribution with density φ^2 . By definition of the change of measure, and

Cauchy-Schwarz, we have that

$$\mathbb{P}_x \left(t^{-1} \sum_{i=1}^{N_t} \varphi(X_t^i) < \delta/\varepsilon | N_t > 0 \right) \leq \frac{\varphi(x)}{t\mathbb{P}_x(N_t > 0)} \mathbb{Q}_x \left[\left(\frac{1}{\sum_{i=1}^{N_t} \varphi(X_t^i)} \right)^2 \right]^{1/2} \mathbb{Q}_x \left(\frac{\sum_{i=1}^{N_t} \varphi(X_t^i)}{t} \leq \delta/\varepsilon \right)^{1/2} \quad (5.10)$$

which is in turn less than

$$\frac{\varphi(x)}{t\mathbb{P}_x(N_t > 0)} \overline{\mathbb{Q}}_x [1/\varphi(\xi_t)^2]^{1/2} \overline{\mathbb{Q}}_x \left(\mathbb{Q}_{\xi_u} \left(\frac{\sum_{i=1}^{N_{t-u}} \varphi(X_{t-u}^i)}{t} \leq \delta/\varepsilon \right) \right)^{1/2}$$

where ξ_s is the position of the spine particle at time s , under $\overline{\mathbb{Q}}_x$. The second inequality holds simply by positivity of φ , and the fact that $\mathbb{Q}_x = \overline{\mathbb{Q}}_x$ on \mathcal{F}_t , the filtration generated by the process, but not the position of the spine, up to time t . Now, the first two terms in the product (5.10) do not depend on δ , and are uniformly bounded in t and x , for $t > 1$ say. This follows from Remark 4.10 (uniform asymptotic for the survival probability), and Lemma 4.3 (convergence of the spine.) The final term of the product is also, by Lemma 4.3 again, less than or equal to

$$\left(C e^{-\gamma u} + \int_D \varphi(y)^2 \mathbb{Q}_y \left(\frac{\sum_{i=1}^{N_{t-u}} \varphi(X_{t-u}^i)}{t} \leq \delta/\varepsilon \right) dy \right)^{1/2} \quad (5.11)$$

for all $u > 1$ say, where C does not depend on x, u or t . By changing measure back to \mathbb{P}_y in the integrand, and again applying Cauchy-Schwarz, we see that the second term in the square root in (5.11) above is bounded by

$$\int_D \varphi(y)^2 \left(\frac{\mathbb{P}_y(N_{t-u} > 0) \mathbb{E}_y[(\sum_{i=1}^{N_{t-u}} \varphi(X_{t-u}^i))^2 | N_{t-u} > 0]^{1/2}}{\varphi(y)} \mathbb{P}_y \left(\frac{\sum_{i=1}^{N_{t-u}} \varphi(X_{t-u}^i)}{t-u} \leq \frac{t}{t-u} \delta/\varepsilon \middle| N_{t-u} > 0 \right)^{1/2} \right) dy.$$

However we know that $\varphi(y)^{-1} \mathbb{P}_y(N_{t-u} > 0) \mathbb{E}_y[(\sum_{i=1}^{N_{t-u}} \varphi(X_{t-u}^i))^2 | N_{t-u} > 0]^{1/2}$ is uniformly bounded in y as long as $t-u > 1$ say (using the moments we calculated in Lemma 5.1 and the asymptotic for the survival probability). Since everything is positive we can take this bound outside of the integral. Furthermore, by Remark 5.2 we know that if we let $g'(\delta)$ be the square root of the probability that an exponential random variable with mean $\lambda \int_D \varphi(y)^3 dy$ is less than δ/ε , then $g'(\delta) \downarrow 0$ as $\delta \downarrow 0$, and

$$\int_D \varphi(y)^2 \mathbb{P}_y \left(\frac{\sum_{i=1}^{N_{t-u}} \varphi(X_{t-u}^i)}{t-u} \leq \frac{t}{t-u} \delta/\varepsilon \middle| N_{t-u} > 0 \right)^{1/2} dy \longrightarrow g'(\delta)$$

as $t \rightarrow \infty$, for any fixed u . Putting all of this together proves the result. \square

We will use this to prove a stronger version of Corollary 1.7. We know by the Corollary that the average value of $f(v)$ among all vertices v at large height in one tree, given survival, converges to $(f, \varphi) / (1, \varphi)$. The next Lemma will tell us that in fact we need only look at the average over a large enough subset of these vertices. This will be helpful to us for the proof of Theorem 1.1.

Lemma 5.4. *Let f be as in Corollary 1.7. Then for any $\varepsilon, \rho > 0$ and $x \in D$*

$$\mathbb{P}_x (B_t(\rho) | N_t > 0) := \mathbb{P}_x \left(\bigcup_{\rho t \leq M \leq N_t} \left\{ \left| \frac{\sum_{i=1}^M f(X_t^i)}{M} - \frac{(f, \varphi)}{(1, \varphi)} \right| > \varepsilon \right\} \middle| N_t > 0 \right) \rightarrow 0$$

as $t \rightarrow \infty$.

Proof. We will prove the Lemma by looking at the tree conditioned to survive until time t , and dividing the set of vertices at time t into families, depending on whether or not they have the same ancestor at some earlier time. This earlier time will be chosen such that with high probability, the average value of f over the positions of any one of these families is close to $(f, \varphi) / (1, \varphi)$. This will show that at many places along the t th generation (vertices at height t) the average of f , taken over the positions of all previous vertices at this height, is close to what we want. To extend this to all vertices far enough along the t th generation, we will show that the size of each of these families is very small compared to t .

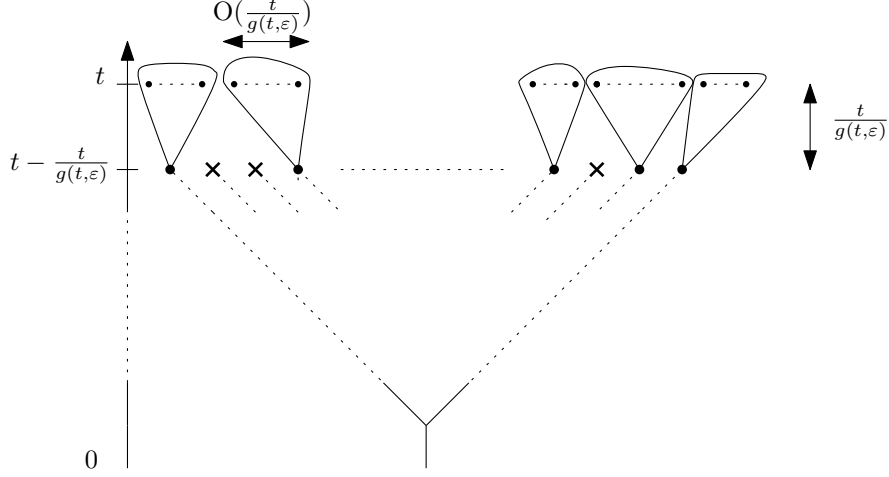


Figure 5.1: Sketch of the argument. There are $O(g(t, \varepsilon))$ particles at time $t - t/g(t, \varepsilon)$ with descendants at time t (marked with dots). Each of these families is likely to be *good* and has size $O(t/g(t, \varepsilon))$.

To do this, fix $\varepsilon > 0$ and write

$$p(t, \varepsilon) := \sup_{x \in D} \mathbb{P}_x \left(\left| \frac{\sum_{i=1}^{N_t} f(X_t^i)}{N_t} - \frac{(f, \varphi)}{(1, \varphi)} \right| > \varepsilon/2 \mid N_t > 0 \right),$$

which by Corollary 5.3, converges to 0 as $t \rightarrow \infty$. This means that we can choose a function $g(t, \varepsilon)$ such that $g(t, \varepsilon) \rightarrow \infty$ as $t \rightarrow \infty$, but

$$g(t, \varepsilon) p \left(\frac{t}{g(t, \varepsilon)}, \varepsilon \right) \rightarrow 0 \quad (5.12)$$

as $t \rightarrow \infty$. Indeed, since $\sup\{p(u, \varepsilon); u \geq t\}$ converges monotonically to 0 as $t \rightarrow \infty$, you can choose $g(t, \varepsilon)$ less than \sqrt{t} but still converging to ∞ , such that $g(t, \varepsilon) \sup\{p(u, \varepsilon); u \geq \sqrt{t}\} \rightarrow 0$ as $t \rightarrow \infty$. Then since $p(t/g(t, \varepsilon), \varepsilon) \leq \sup\{p(u, \varepsilon); u \geq t/g(t, \varepsilon)\} \leq \sup\{p(u, \varepsilon); u \geq \sqrt{t}\}$, the function g will satisfy (5.12).

As mentioned above, we will break up the particles at time t into families. Two vertices will be in the same family if they have a common ancestor at time $t - t/g(t, \varepsilon)$. We let the number of these families be $\hat{N}_{t-t/g(t, \varepsilon)}$ and set m_i to be the average value of f among the i th family. Here the order of the families corresponds to the order of the ancestors at time $t - t/g(t, \varepsilon)$. The key to the proof of this Lemma will be to show that

$$\mathbb{P}_x \left(\bigcup_{i=1}^{\hat{N}_{t-t/g(t, \varepsilon)}} \left\{ \left| m_i - \frac{(f, \varphi)}{(1, \varphi)} \right| > \varepsilon/2 \right\} \mid N_t > 0 \right) \rightarrow 0 \quad (5.13)$$

as $t \rightarrow \infty$. The reason for this is that there are order $g(t, \varepsilon)$ particles that have descendants at time t , and the probability that they are bad in the sense of (5.13) is less than $p(t/g(t, \varepsilon), \varepsilon)$ by definition. Then (5.12) provides the result.

To prove this rigorously however, it is more convenient to consider the unconditioned version of the probability in (5.13), noting that the event clearly does not occur if $N_t = 0$. To analyse this probability, we condition on the

total collection of particles at time $t - t/g(t, \varepsilon)$. There are order 1 of these, uniformly in t by Markov's inequality (i.e. for any $\delta > 0$ there exists a K such that $\mathbb{P}_x(N_{t-t/g(t, \varepsilon)} > K) \leq \delta$ for all t .) Moreover, for any one of them, the probability that it

(a) has a descendant at time t , and

(b) the average value of f over all its descendants at time t is more than $\varepsilon/2$ away from $(f, \varphi)/(1, \varphi)$

is less than some constant times $\frac{g(t, \varepsilon)}{t} p(t/g(t, \varepsilon), \varepsilon)$. This follows from the definition of p and the asymptotic for the survival probability. Multiplying by t to account for the conditioning in (5.13) and applying (5.12) gives (5.13).

The upshot of (5.13) is that we now know, letting σ_i be the number of particles in the i th family at time t , that

$$\mathbb{P}_x(A_t | N_t > 0) := \mathbb{P}_x \left(\bigcup_{i=1}^{\hat{N}_{t-t/g(t, \varepsilon)}} \left\{ \left| \frac{\sum_{j=1}^{\sigma_1 + \dots + \sigma_i} f(X_t^j)}{\sigma_1 + \dots + \sigma_i} - \frac{(f, \varphi)}{(1, \varphi)} \right| > \varepsilon/2 \right\} \middle| N_t > 0 \right) \rightarrow 0 \quad (5.14)$$

as $t \rightarrow \infty$ (observe that the families are clearly grouped together in the ordering of generation t .) This tells us that we have many “good” vertices along generation t , where the average value of f (considered over previously visited vertices) is close to what we want. To complete the proof, we must show that gaps between these good vertices, i.e. the lengths σ_i , are not too long.

To do this, we will prove that

$$\mathbb{P}_x(A'_t | N_t > 0) := \mathbb{P}_x \left(\bigcup_{i=1}^{\hat{N}_{t-t/g(t, \varepsilon)}} \left\{ \sigma_i > \frac{t}{(g(t, \varepsilon))^{1/3}} \right\} \middle| N_t > 0 \right) \rightarrow 0 \quad (5.15)$$

as $t \rightarrow \infty$. For this we apply a similar argument to above. We consider the unconditioned probability and condition on the system at time $t - t/g(t, \varepsilon)$. By Markov's inequality there are order 1 particles at this time, uniformly in t , and for any one of them the probability that it

(a) has a descendant at time t , and

(b) the total number of descendants at time t is greater than $t/(g(t, \varepsilon))^{1/3}$

is less than some constant times

$$\frac{g(t, \varepsilon)}{t} \times g(t, \varepsilon)^{-4/3}.$$

The first term in the product comes from the asymptotic for the survival probability. The second comes from the fact that, given survival of a process to time $t/g(t, \varepsilon)$, the total number of particles is roughly $t/g(t, \varepsilon)$ times an exponential random variable by Theorem 1.6. We use Markov's inequality to get the explicit bound (uniformly in the starting point). Again multiplying by t to account for the conditioning gives (5.15).

Let us now show that $\mathbb{P}_x(B_t(\rho) | N_t > 0) \rightarrow 0$ as $t \rightarrow \infty$. By the above work, and a union bound it is enough to show that

$$B_t(\rho) \subset A_t \cup A'_t$$

for all t large enough. Suppose we are on the event $\{A_t \cup A'_t\}^c$, and for every $\rho t \leq M \leq N_t$, set $k(M) = \sigma_1 + \dots + \sigma_i$, where i is such that $\sigma_1 + \dots + \sigma_i \leq M \leq \sigma_1 + \dots + \sigma_{i+1}$. Then

$$\left| \frac{\sum_{i=1}^{k(M)} f(X_t^i)}{k(M)} - \frac{(f, \varphi)}{(1, \varphi)} \right| \leq \varepsilon/2$$

for all $\rho t \leq M \leq N_t$ simultaneously. However, since we are on the event $\{A'_t\}^c$ we also have

$$\left| \frac{\sum_{i=1}^M f(X_t^i)}{M} - \frac{\sum_{i=1}^{k(M)} f(X_t^i)}{k(M)} \right| \leq 4 \frac{(f, \varphi)}{(1, \varphi)} \frac{1}{\rho(g(t, \varepsilon))^{1/3}}. \quad (5.16)$$

Thus, for all t large enough, we must also be on the event $\{B_t(\rho)\}^c$.

□

Remark 5.5. *The proof of Theorem 1.6 directly extends to the case of a general branching diffusion, by applying a generalised version of the Many-to-Few Lemma, see [HR15, Lemma 1]. Then Corollary 1.7 and Lemma 5.4 follow as above.*

6 Convergence to the Brownian CRT

From this point onwards, we will assume that $\varphi \in C^1(\overline{D})$ as in the statement of Theorem 1.1. In particular this means that $|\nabla\varphi|$ is bounded on D .

Recall from Section 2.1 that we can consider our branching diffusions as continuous planar trees, where all the vertices are marked with a position. We write \mathcal{P}_x for the law of a sequence of i.i.d critical branching Brownian motion trees, each starting at $x \in D$. We will explore these trees in a *depth-first* order. This exploration is defined as follows:

- We start at the root of the first tree and move upwards (i.e. increasing height) at speed one. Whenever we reach a branching point we take the left branch.
- When we can no longer continue, we jump instantaneously to the most recent branching point that we have visited.
- We then repeat the process, starting to explore along the right branch emanating from this point.
- Whenever we can no longer continue, we jump instantaneously to the most recent branch point that we have visited, but not already jumped up to.
- When we reach the end of the first tree, we jump instantaneously to the root of the next tree and repeat.

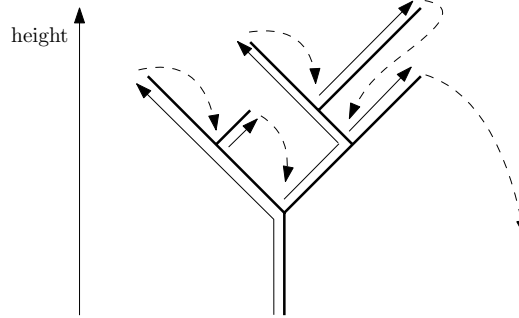


Figure 6.1: A sketch of the depth-first exploration of a continuous tree. Full arrowed lines represent motion at speed one in the vertical direction. Dotted arrowed lines represent instantaneous jumps.

This is the analogue of the *lexicographical* ordering for discrete trees. Recording the height of vertices as we traverse the trees in this way gives us the *height process* H_t associated with the sequence. To show the convergence in Theorem 1.1 it will be important to show that this height process, when rescaled appropriately, looks like a reflected Brownian motion. To do this, we introduce a further process, S_t , which will turn out to be a martingale.

In the following, we will say that a vertex in the sequence of trees has been *visited* by time t if the exploration has passed through that point before time t . We will say that a vertex, that is also a branch point, has been *explored* by time t if it has been visited *and* jumped back to before time t . Recall that the mark associated with a vertex v corresponds to the spatial position of the particle it represents: we denote this by v^* . Finally, we write $Y(v)$ for the set of branch points that have been visited but not explored before the time that v is visited.

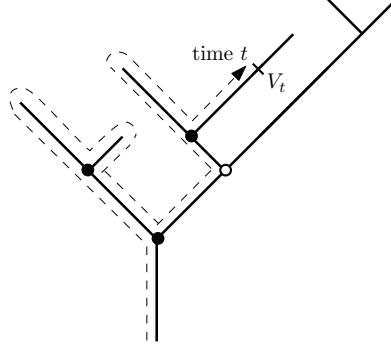


Figure 6.2: The exploration up to time t . Branch points with filled circles have been visited and explored before time t . Branch points with empty circles have been visited but not explored. Note that the branch point furthest to the right has neither been visited nor explored before time t . $Y(V_t)$ is the set of branch points with empty circles.

Definition 6.1. Let V_t be the vertex that is visited at time t in the depth-first exploration, writing V_t^* as usual for its position in D . We define for $t \geq 0$

$$S_t = \varphi(V_t^*) + \sum_{v \in Y(V_t)} \varphi(v^*) - \Lambda_t \varphi(x).$$

Here, Λ_t is the index of the tree being visited at time t .

In fact, S_t is very closely related to the martingale M_t . Essentially they are the same process, but explored in different orders, and we will see that this is enough to preserve the martingale property. We would like to approximate the height process by S_t , and then apply an invariance principle for the martingale. This is an analogous idea to that used to prove convergence of Galton-Watson processes to the CRT in [GD02], where S_t here plays the role of the *Lukasiewicz path*. We first record a property of this process, which will be essential to showing a relationship with the height function:

Lemma 6.2. Let $S'_t = S_t - \varphi(V_t^*)$ and $I'_t = \inf_{0 \leq s \leq t} S'_s$. Then

$$S'_t - I'_t = \sum_{v \in Y(V_t)} \varphi(v^*)$$

Proof. Since φ is positive, it is clear that $I'_t = -\Lambda_t \varphi(x)$. This implies the result. \square

Definition 6.3. The process on the right hand side in Lemma 6.2 also makes sense for a depth-first exploration of a single tree, and we denote such a process by \hat{S} . Note that this is a process that starts at 0 and is positive until the exploration of the tree is finished, i.e. an excursion.

We also write \hat{S}^y for \hat{S} conditioned to reach level y . Then \hat{S}^y is equal in law to $S'_t - I'_t$, restricted to the first excursion in which it exceeds y .

Observe that we can decompose S into continuous and discontinuous parts, S^c and S^d . Indeed, if we let S^d be the pure jump process that jumps up by $\varphi(v^*)$ whenever the exploration reaches a branch point vertex v , then $S^c = S - S^d$ is continuous. In fact, S^c is simply $\varphi(V_t^*)$ minus a compensating sum that makes it continuous. If the exploration reaches the end of a branch at time t , then $\lim_{s \uparrow t} \varphi(V_s^*) = 0$, but $\lim_{s \downarrow t} \varphi(V_s^*) > 0$. It is easy to verify that S^c is just $\varphi(V_t^*)$ with the jumps subtracted whenever they occur.

Remark 6.4. For a general diffusion and branching mechanism, as in the statement of Theorem 1.1, we define the process S in almost the same way, setting φ to be the first eigenfunction of the generator as usual. In this case we will only say that a branch point has been explored when all of the subtrees branching from the point have been partially explored. If a branch point has been visited but not explored before time t we let $k_t(v)$ be the number of

these subtrees that have not been explored at all before time t . We replace the sum $\sum_{v \in Y(V_t)} \varphi(v^*)$ in the definition of S by $\sum_{v \in Y(v_t)} k_t(v) \varphi(v^*)$. Lemma 6.2 then holds with this sum on the right hand side. When we decompose S into its continuous and discontinuous parts, the jumps are given by $(k - 1)\varphi(v^*)$ whenever a branch point v with k branches is reached. The continuous part is again just $\varphi(V_t^*)$ with the jumps removed.

We will return to this later, but let us now prove an invariance principle for S_t .

6.1 Martingale Convergence

Lemma 6.5. *Under \mathcal{P}_x , for any $x \in D$, $(S_t)_{t \geq 0}$ is a locally square-integrable martingale with respect to the natural filtration generated by the depth-first exploration. Its predictable quadratic variation is given by*

$$\langle S \rangle_t = \int_0^t \lambda \varphi(V_s^*)^2 + |\nabla \varphi(V_s^*)|^2 ds.$$

Proof. First observe that breaking up $S = S^d + S^c$ gives us the expression for the predictable quadratic variation. The first term comes from the discontinuous part, which we know jumps, at rate λ , by φ applied to the position of the vertex being visited. The second comes from the fact that we can write $dS_t^c = \varphi(V_t^*)dB_t$ for B a standard Brownian motion. This follows from the description of S_t^c , and the fact that the increments of $(V_t^*)_{t \geq 0}$ are equal in law to those of a Brownian motion.

To see that S is a martingale, we condition on the depth first exploration up to time s and notice that we can write

$$\mathcal{E}_x[S_t - S_s | \mathcal{F}_s] = \mathbb{E}_{V_s^*}[\hat{S}_\tau - \varphi(V_s^*)] + \sum_{v \in Y(V_s)} \mathbb{E}_{v^*}[\hat{S}_{v, \tau_v} - \varphi(v^*)] + \sum_{i=1}^{\infty} \mathbb{E}_x[\hat{S}_{i, \tau_i} - \varphi(x)]$$

by the Markov property, where \hat{S}_τ , $(\hat{S}_{v, \tau_v})_{v \in Y(V_s)}$ and $(\hat{S}_{i, \tau_i})_{i \geq 1}$ are the processes \hat{S} for the subtrees rooted at v_s , $\{v\}_{v \in Y(V_s)}$ and $\{v_i\}_{i \geq 1}$ (the roots of the remaining sequence of trees) respectively, each run up to a stopping time that *does not depend on that subtree* by conditional independence. Thus, to prove the martingale property it is enough to show that

$$\mathbb{E}_x[\hat{S}_t] = \varphi(x)$$

for any $x \in D$ and $t \geq 0$.

To do this, we will approximate \hat{S} by a discrete version $(\hat{S}_n^\delta, n \in \mathbb{N})$, with $\delta \downarrow 0$. This process will be defined by discretising the tree using steps of size δ in the natural way. This results in a discrete tree where every vertex is marked with a position - corresponding to its spatial position in the original tree. To define \hat{S}^δ , set $\hat{S}_0^\delta = \varphi(x)$ (where x is the starting position) and traverse the discrete tree in a depth-first, or lexicographical, order. If you are visiting a vertex with position y at step n , and it has children with positions $(z_j)_{1 \leq j \leq J}$, then set $\hat{S}_{n+1}^\delta - \hat{S}_n^\delta = \sum_{j=1}^J \varphi(z_j) - \varphi(y)$. By considering the martingale M , it is clear that $\mathbb{E}_x[\hat{S}_n^\delta] = \varphi(x)$ for every n and δ . Moreover, for fixed $t \geq 0$ we have that $\hat{S}_{\lfloor t/\delta \rfloor}^\delta \rightarrow \hat{S}_t$ almost surely as $\delta \rightarrow 0$. This is clear after noting that:

- Almost surely the discrete tree only captures single branching events in any step, for all δ small enough.
- If vertex being visited at time $\lfloor t/\delta \rfloor$ in the discrete process corresponds to the vertex at time u in the continuous process, then $|u - t| \leq \delta(1 + W_t)$, where W_t is the total number of deaths plus branch points before time t in the exploration.
- The particle motion is almost-surely continuous.

Since the $\hat{S}_{\lfloor t/\delta \rfloor}^\delta$ are dominated, for example by $\|\varphi\|_\infty$ times the number of branch points in the continuous tree up to level t , the result follows by dominated convergence. \square

Remark 6.6. If we have a generator $L = -\frac{1}{2} \sum_{i,j} a^{ij} \partial_{x_i x_j} + \sum_i b^i \partial_{x_i}$ as in Theorem 1.1 and an offspring distribution A with finite variance, then the same argument implies that S is a martingale. Letting λ, φ be the first eigenvalue/eigenfunction pair for $-L$ we can calculate that the predictable quadratic variation of S is given by

$$\langle S \rangle_t = \int_0^t \frac{\lambda}{m-1} \mathbb{E}[A^2 - A] \varphi(V_s^*)^2 - \lambda \varphi(V_s^*)^2 + \sum_{i,j} \frac{\partial \varphi}{\partial x_i} \frac{\partial \varphi}{\partial x_j} a^{i,j}(V_s^*) ds.$$

The second term here comes from the single particle motion under L and the first comes from the jumps (recall the critical branching rate in this case is $\lambda/(m-1)$).

Proposition 6.7. Let

$$\sigma^2 = \frac{2\lambda \int_D \varphi(y)^3 dy}{(1, \varphi)}.$$

Then

$$\left(\frac{S_{nt}}{\sqrt{n}} \right)_{t \geq 0} \rightarrow (\sigma B_t)_{t \geq 0}$$

in distribution as $n \rightarrow \infty$, with respect to the Skorohod topology.

Proof. Since the jumps of S are bounded, this follows from the functional central limit theorem for martingales [JS87, Theorem 3.22, Chapter VIII] once we can show that for all $t \geq 0$

$$\langle S^n \rangle_t \rightarrow \sigma t$$

in probability as $n \rightarrow \infty$. Here $(S_t^n)_{t \geq 0} = (S_{nt}/\sqrt{n})_{t \geq 0}$. However, we can write

$$\frac{\langle S^n \rangle_t}{t} = \frac{1}{nt} \int_0^{nt} \lambda \varphi(V_s^*)^2 + |\nabla \varphi(V_s^*)|^2 ds.$$

Then, since φ and $|\nabla \varphi|$ are bounded, this follows immediately from Proposition 6.8 below. We recover σ^2 by integrating by parts to see that $(\lambda \varphi(\cdot)^2 + |\nabla \varphi(\cdot)|^2, \varphi) = 2\lambda \int_D \varphi(y)^3 dy$. \square

Proposition 6.8. Suppose that f is a bounded, measurable function and let $(V_s^*)_{1 \leq s \leq t}$ be the positions of vertices visited in the depth-first exploration before time t as usual. Then

$$Q_t := \frac{1}{t} \int_0^t f(V_s^*) ds \rightarrow \frac{(f, \varphi)}{(1, \varphi)}$$

in \mathcal{P}_x -probability as $t \rightarrow \infty$.

Remark 6.9. In the general case (assuming Proposition 6.8) we can integrate by parts, using that φ is an eigenvector of L and L is self-adjoint, to see that $\langle S^n \rangle_t \rightarrow \sigma t$ as $n \rightarrow \infty$, where

$$\sigma^2 = \frac{\lambda \mathbb{E}[A^2 - A] \int_D \varphi(y)^3 dy}{(m-1)(1, \varphi)}.$$

Since the jumps are not necessarily bounded in this case, to apply the FCLT for martingales we need to verify a Lindenberg-Feller type condition (for example, [JS87, (3.23), Theorem 3.22, Chapter VIII].) However, this follows immediately from the fact that we are assuming the offspring distribution to have finite variance. Then Proposition 6.7 holds.

Before we prove Proposition 6.8, let us record some of the consequences of Proposition 6.7.

Proposition 6.10. *Let Λ_t be the index of the tree being visited in the exploration at time t (so $\Lambda_0 = 1$). Then we have the joint convergence*

$$\left(\frac{S'_{nt} - I'_{nt}}{\sqrt{n}}, \frac{\Lambda_{nt}}{\sqrt{n}} \right)_{t \geq 0} \rightarrow \left(\sigma |\beta_t|, \frac{\sigma}{\varphi(x)} L_t^0(\beta) \right)_{t \geq 0} \quad (6.1)$$

as $n \rightarrow \infty$, in distribution with respect to the Skorohod topology. Here, β is a standard Brownian motion started at 0 and $L_t^0(\beta)$ is the local time of β at 0. Furthermore, for any $y > 0$, we have

$$\left(\frac{\hat{S}_{nt}^{y\sqrt{n}}}{\sqrt{n}} \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{(d)} \left(\sigma e_t^{y/\sigma} \right)_{t \geq 0} \quad (6.2)$$

where $e^{y/\sigma}$ is a Brownian excursion conditioned to reach height y/σ .

Proof. From the definition of S' , we have that $|S'_{nt}/\sqrt{n} - S_{nt}/\sqrt{n}| \leq \|\varphi\|_\infty/\sqrt{n}$ for all $t \geq 0$ and so Proposition 6.7 implies that

$$\left(\frac{S'_{nt}}{\sqrt{n}} \right)_{t \geq 0} \rightarrow (\sigma B_t)_{t \geq 0}$$

as $n \rightarrow \infty$ as well. Writing $\underline{B}_t = \inf_{0 \leq s \leq t} B_s$ this implies the joint convergence

$$\left(\frac{S'_{nt} - I'_{nt}}{\sqrt{n}}, -\frac{I'_{nt}}{\sqrt{n}} \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{(d)} (\sigma(B_t - \underline{B}_t), -\sigma \underline{B}_t)_{t \geq 0} \quad (6.3)$$

where the right-hand side by Lévy's theorem, see for example [RY91, Chapter VI, Theorem VI.2.3], is equal in distribution to

$$(\sigma |\beta_t|, \sigma L_t^0(\beta))_{t \geq 0}.$$

However, we know that $\Lambda_t = -I'_t/\varphi(x)$, and so (6.1) follows. For the second claim of the Proposition, we follow [GD02, Proposition 2.5.2]. It is well known that you can construct the process $e^{y/\sigma}$ from a standard Brownian motion β by taking

$$e_t^{y/\sigma} = |\beta_{(G+t) \wedge D}|$$

where $T = \inf\{t \geq 0 : |\beta_t| \geq y/\sigma\}$, $G = \sup\{t \leq T : \beta_t = 0\}$ and $D = \inf\{t \geq T : \beta_t = 0\}$. By the Skorohod representation theorem and (6.1) we also know that there exists a process

$$(Z_t^{(n)}, \Lambda_t^{(n)})_{t \geq 0} \xrightarrow{(d)} \left(\frac{S'_{nt} - I'_{nt}}{\sqrt{n}}, \frac{\Lambda_{nt}}{\sqrt{n}} \right)_{t \geq 0}$$

such that

$$(Z_t^{(n)}, \Lambda_t^{(n)})_{t \geq 0} \xrightarrow[n \rightarrow \infty]{} \left(\sigma |\beta_t|, \frac{\sigma}{\varphi(x)} L_t^0(\beta) \right)_{t \geq 0}$$

uniformly on every compact set almost surely. This is because Skorohod convergence is equivalent to local uniform convergence when the limit is continuous. Define $T^{(n)} = \inf\{t \geq 0 : Z_t^{(n)} \geq y\}$ for this sequence of processes, and $G^{(n)}, D^{(n)}$ in the same way as G and D above. By the remark in Definition 6.3 and (6.1), if we can prove that $G^{(n)} \rightarrow G$ and $D^{(n)} \rightarrow D$ almost surely, we will be done. First note that since β must exceed x/σ immediately after time T , we have that $T^{(n)} \rightarrow T$ almost surely. This implies straight away that for all $t < D$ we have $t \leq D^{(n)}$ for all n large enough almost surely. Now we must show that for all $t > D$ we have $t \geq D^{(n)}$ for all n large enough almost surely. These facts together (along with the corresponding results for G) are enough to prove the convergence. To see the final claim, we use the convergence of the local time. For any $t > D$, we have using basic properties of Brownian local time that $L_t^0 > L_D^0 = L_T^0$. The convergence of the local time therefore tells us that $\Lambda_t^{(n)} > \frac{\sigma}{\varphi(x)} L_T^0$ for all n large enough almost surely. Since

$$\Lambda_{T^{(n)}}^{(n)} \rightarrow \frac{\sigma}{\varphi(x)} L_T^0$$

almost surely, this implies that we have also have $\Lambda_t^{(n)} > \Lambda_{T^{(n)}}^{(n)}$ for all n large enough almost surely. Using the fact that $\Lambda^{(n)}$ stays constant on $[T^{(n)}, D^{(n)})$, we see that $t \geq D^{(n)}$. \square

Remark 6.11. *The above proof is not affected by changing the generator or offspring distribution, since it relies only on the convergence from Proposition 6.7.*

The rest of this subsection will be devoted to the proof of Proposition 6.8, but we will first need some preliminary estimates. We write $h(v)$ for the height of a vertex v in the the exploration.

Lemma 6.12. *Let L be the total length of a branching Brownian motion process (i.e. how long it takes to traverse the tree in depth-first order). Then for all $x \in D$*

$$\mathbb{P}_x(L > t) \gtrsim \frac{\varphi(x)}{\sqrt{t}}.$$

Let U_t be a vertex picked uniformly at random from those visited before time t in the depth-first exploration under \mathcal{P}_x . Write U_t^* as usual for the position in D corresponding to its mark. Finally, denote by $\bar{\mathcal{P}}_x$ the law on the sequence of trees plus U_t . Then we have the following control on the height of the uniform vertex.

Lemma 6.13. *For all $x \in D$*

$$\bar{\mathcal{P}}_x \left(h(U_t) \geq C\sqrt{t} \right) \rightarrow 0$$

as $C \rightarrow \infty$, uniformly in t .

We will now show how we may deduce Proposition 6.8, and then go on to prove the Lemmas. Let N_s^i be the number of particles at level s in the i th tree of our exploration. Also write $(X_s^{i,j})_{1 \leq j \leq N_s^i}$ for the positions of these particles. Finally, let $I(t)$ be the index of the tree in which the uniform vertex lies.

Proof of Proposition 6.8. We first show that $\mathcal{E}_x[Q_t] \rightarrow \frac{(f, \varphi)}{(1, \varphi)}$ as $t \rightarrow \infty$. Let m_t be the average value of f among the vertices at height $h(u_t)$ of the $I(t)$ th tree, that have been visited before time t . Observe that by conditioning on m_t , we have

$$\mathcal{E}_x[Q_t] = \bar{\mathcal{E}}_x[f(U_t^*)] = \bar{\mathcal{E}}_x[m_t]$$

as $t \rightarrow \infty$. This is because, given the positions of these particles, we know that u_t is chosen uniformly from them. We will aim to show that for fixed $\varepsilon > 0$,

$$\bar{\mathcal{P}}_x(A_t) := \bar{\mathcal{P}}_x \left(\left| m_t - \frac{(\varphi, f)}{(\varphi, 1)} \right| > \varepsilon \right) \rightarrow 0 \quad (6.4)$$

as $t \rightarrow \infty$. Then since m_t is bounded, the convergence in expectation will follow.

For the i th tree in our exploration and $s > 0$, let $m_{s,t}^i$ be the average value of f among the vertices of tree i at height s that are visited before time t , and $\tilde{N}_{s,t}^i$ be the number of such vertices. Also write $A_{s,t}^i$ for the event that $|m_{s,t}^i - (\varphi, f) / (\varphi, 1)| > \varepsilon$. Then, conditioning on the entire sequence of trees, we can write our probability as

$$\frac{1}{t} \mathcal{E}_x \left[\int_0^\infty \sum_{i=1}^\infty \mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{1}_{A_{s,t}^i} \tilde{N}_{s,t}^i ds, \right]$$

since the probability of picking a vertex in tree i at height ds is $\tilde{N}_s^i/t ds$. Note that we can interchange the integral, expectation and sum as we like here by Fubini (the expression being bounded by 1.) Now, given $\delta > 0$, by Lemma 6.13 we can choose C such that $\mathbb{P}(h(U_t) \geq C\sqrt{t}) < \frac{\delta}{3}$ for all t . Similarly we can define $R(t) > 0$ such that $\bar{\mathcal{P}}_x(h(U_t) \leq R(t)) = \frac{\delta}{3}$. Note that $R(t) \rightarrow \infty$ as $t \rightarrow \infty$ by the law of large numbers. Indeed, for any $K > 0$ the proportion of time spent below height K in the exploration is less than or equal to $\sum_{i=1}^{\Lambda_t} p_K^i/t$ where p_K^i is the time

spent below level K in the i th tree. We also have that t is greater than or equal to $\sum_{i=1}^{\Lambda_t-1} L^i$ where L^i is the length of the i th tree. However, p_K^i has finite variance and L^i does not, and so the strong law of large numbers allows us to conclude.

Since $\mathbb{1}_{A_i^{s,t}} \leq 1$ this tells us that

$$\overline{\mathcal{P}}_x(m_t) \leq \frac{2\delta}{3} + \frac{1}{t} \mathcal{E}_x \left[\int_{R(t)}^{C\sqrt{t}} \sum_{i=1}^{\infty} \mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{1}_{A_{s,t}^i} \tilde{N}_s^i ds \right] \quad (6.5)$$

where by Fubini we can rewrite the expectation term as

$$\frac{1}{t} \int_{R(t)}^{C\sqrt{t}} \sum_{i=1}^{\infty} \mathcal{E}_x \left[\mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{1}_{A_{s,t}^i} \tilde{N}_s^i \right] ds. \quad (6.6)$$

Now for each i , we condition on \mathcal{G}_{i-1} : the σ -algebra generated by the first $i-1$ trees. Note that \mathcal{G}_{i-1} is independent of the i th tree. Moreover, the event $\{i \leq \Lambda_t\}$ and the start time τ_i of the i th tree, are measurable with respect to \mathcal{G}_{i-1} . Thus we can write

$$\mathcal{E}_x \left[\mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{1}_{A_{s,t}^i} \tilde{N}_s^i \middle| \mathcal{G}_{i-1} \right] = \mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{E}_x \left[\mathbb{1}_{A_s(t-\tau_i)} N_s(t-\tau_i) \middle| N_s > 0 \right] \mathbb{P}_x(N_s > 0)$$

where the expectation is now with respect to a single branching Brownian motion tree, $N_s(r)$ is the number of particles at level s before time r in a depth-first exploration of the tree, and $A_s(r)$ is the event that the average of f among these particles is more than ε away from $(\varphi, f) / (\varphi, 1)$.

We note here, by Theorems 1.4 and 1.6, that there exists a K such that

$$\mathbb{E}_x \left[N_s^2 \middle| N_s > 0 \right]^{1/2} \leq Ks, \quad \mathbb{P}_x(N_s > 0) \leq K/s \quad (6.7)$$

for all s . We can also, by Lemma 6.12, choose this K such that

$$\mathcal{E}_x[\Lambda_t] \leq K\sqrt{t} \quad (6.8)$$

for all t . Indeed, $\mathcal{E}_x[\Lambda_t]$ is less than the expectation of geometric random variable, whose success rate is $\mathbb{P}_x(L > t) \gtrsim 1/\sqrt{t}$. Decomposing on whether or not $N_s(t-\tau_i)$ is bigger than $s\delta/6CK^2$ we have

$$\begin{aligned} \mathbb{E}_x \left[\mathbb{1}_{A_s(t-\tau_i)} N_s(t-\tau_i) \middle| N_s > 0 \right] &\leq \frac{\delta s}{6CK^2} + \mathbb{E}_x \left[\mathbb{1}_{B_s(\delta/6CK^2)} N_s(t-\tau_i) \middle| N_s > 0 \right] \\ &\leq \frac{\delta s}{6CK^2} + \mathbb{E}_x \left[N_s^2 \middle| N_s > 0 \right]^{1/2} \mathbb{P}_x \left(B_s(\delta/6CK^2) \middle| N_s > 0 \right)^{1/2} \end{aligned}$$

where $B_s(\cdot)$ is the event from Lemma 5.4. Using (6.7), and the conditioning above, we have that

$$\mathcal{E}_x \left[\mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{1}_{A_{s,t}^i} \tilde{N}_s^i \right] \leq \left(\frac{\delta}{6CK} + K^2 \mathbb{P}_x \left(B_s(\delta/6CK^2) \middle| N_s > 0 \right)^{1/2} \right) \mathcal{E}_x[\{i \leq \Lambda_t\}]. \quad (6.9)$$

This means that the expression in (6.6) is less than

$$\left(\frac{\delta}{6CK} + K^2 \sup_{s \geq R(t)} \left\{ \mathbb{P}_x \left(B_s(\delta/6CK^2) \middle| N_s > 0 \right)^{1/2} \right\} \right) \frac{1}{t} \int_{R(t)}^{C\sqrt{t}} \sum_{i=1}^{\infty} \mathcal{E}_x \left[\mathbb{1}_{\{i \leq \Lambda_t\}} \right].$$

However, by Fubini and (6.8), this is less than

$$\frac{\delta}{6} + CK^3 \sup_{s \geq R(t)} \left\{ \mathbb{P}_x \left(B_s(\delta/6CK^2) \middle| N_s > 0 \right)^{1/2} \right\}.$$

Using Lemma 5.4, and the fact that $R(t) \rightarrow \infty$, we see that this is less than $\frac{\delta}{3}$ for all t large enough. Substituting in to (6.5) proves (6.4).

To complete the proof of the Proposition, we must also show that

$$\mathcal{E}_x[Q_t^2] \rightarrow \frac{(f, \varphi)^2}{(1, \varphi)^2}$$

as $t \rightarrow \infty$. However, letting U_t^1 and U_t^2 be two vertices picked independently uniformly from those visited before time t (again denoting the extended law by $\bar{\mathcal{E}}_x$) we see that

$$\mathcal{E}_x[Q_t^2] = \bar{\mathcal{E}}_x[f(U_t^{1*})f(U_t^{2*})].$$

Let m_t^i and A_t^i correspond to m_t and A_t for U_t^i , $i = 1, 2$. Then the same reasoning as above tells us that

$$\bar{\mathcal{E}}_x[f(U_t^{1*})f(U_t^{2*})] = \bar{\mathcal{E}}_x[m_t^1 m_t^2]$$

and we also have, by a union bound, that

$$\bar{\mathcal{P}}_x[A_t^1 \cup A_t^2] \rightarrow 0$$

as $t \rightarrow \infty$. The result follows in exactly the same way. \square

Proof of Lemma 6.12. We will prove this result using our asymptotic for the survival probability. Since we have

$$\mathbb{P}_x(L > t) \geq \mathbb{P}_x(L > t \mid N_{\sqrt{at}} > 0) \mathbb{P}_x(N_{\sqrt{at}} > 0) \quad (6.10)$$

and we know that the probability of survival until time \sqrt{at} decays like $\varphi(x)/\sqrt{at}$, it is enough to show there exists an a such that $\mathbb{P}_x(L > t \mid N_{\sqrt{at}} > 0)$ is bounded below, uniformly in t . In fact, we will prove a slightly stronger statement, as it will also be of use later on. We will show that

$$\mathbb{P}_x(L \leq cs^2 \mid N_s > 0) \rightarrow 0 \quad (6.11)$$

as $c \rightarrow 0$, uniformly in s . Clearly this is enough to prove the Lemma.

To do this, we first fix some compact subdomain $D' \subset D$. Then, we condition on the positions X_s^i of the particles at time s that lie within D' , and consider the subprocesses continuing from these points. In particular, we consider the contributions that each subprocess makes to L from its first μs generations. Note that, given the positions X_s^i , these are independent random variables, with distribution equal to that of $\int_0^{\mu s} N_u du$ for a branching Brownian motion started at X_s^i . This means that they all have mean $\geq C\mu s$ and variance $\leq C'\mu^3 s^3$ for some fixed $C, C' > 0$. The statement concerning the mean follows from the expression for $\mathbb{E}_y[N_u]$, which is bounded below by Lemma 5.1 for all y in the subdomain and all $u > 0$ by some C . For the variance, note that for $r < u$

$$\mathbb{E}_y[N_r N_u] = \mathbb{E}_y[N_r \mathbb{E}_x[N_u \mid \mathcal{F}_r]] \leq C'' \mathbb{E}_y[N_r^2]$$

for some $C'' > 0$ (not depending on r, u or y). Since $[\mathbb{E}_y[N_r^2]] \lesssim r$ uniformly in y , the claim follows by integration.

By Theorem 1.6 we also know that for any $\delta > 0$ there exist m and S , such that the probability of having less than ms vertices at time s in D' (conditioned on survival) is less than δ for all $s \geq S$. On the complementary event, conditioned on the tree up to time s , we have ms independent random variables with mean $\geq C\mu s$ and variance $\leq C'\mu^3 s^3$. By the standard application of Markov's inequality to sums of independent random variables, the probability that their sum is more than $m\mu Cs^2/2$ away from its mean is less than $4C'\mu/C^2 m$ for all $s \geq S$. Since the mean is greater than $m\mu Cs^2$, this has to occur for the sum to be less than $m\mu Cs^2/2$. Taking μ to 0, and noting that L is greater than this sum, shows that the left hand side of (6.11) can be made less than δ as long as c is small enough and s is large enough. In fact, we can choose c small enough that this will hold for all s , due to the continuity in s of the probability in (6.11) (which you can prove by dominated convergence.) Thus the claim is proved. \square

Proof of Lemma 6.13. As explained in the proof of Proposition 6.8, Lemma 6.12 immediately implies that $\mathcal{E}_x[\Lambda_t] \lesssim \sqrt{t}$. Then using our asymptotic for the survival probability, we have

$$\bar{\mathcal{P}}_x \left(h(U_t) \geq C\sqrt{t} \right) \leq \mathcal{E}_x \left[\sum_{i=1}^{\infty} \mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{1}_{\{N_{C\sqrt{t}}^i > 0\}} \right] \lesssim \frac{1}{C\sqrt{t}} \mathcal{E}_x[\Lambda_t]$$

where the implied constant does not depend on C, x, t . □

Remark 6.14. The proofs of Proposition 6.8 and Lemmas 6.12 and 6.13 do not need any adaptation for the more general set up.

6.2 Connection with the height function

In order to make use of the above invariance principle, we must connect the martingale with the height function of our trees. In this section we will look at the height function, and the process \hat{S} from Definition 6.3, for the depth-first exploration of a single tree conditioned to be large. We know by Lemma 6.2 that the value of the \hat{S} when it visits a vertex v in the tree is equal to

$$\hat{S}(v) := \sum_{u \in Y(v)} \varphi(u^*).$$

We will show that for vertices with large heights, this sum is close to a constant times the height. Our approach will use an ergodicity property for the spine particle in the system under $\bar{\mathbb{Q}}_x$, and is inspired from [HR14].

In the following, given $\eta > 0$, we will say that a vertex v in a branching Brownian motion tree is η -bad if

$$\left| \frac{\hat{S}(v)}{h(v)} - \lambda \int_D \varphi(y)^3 dy \right| > \eta. \quad (6.12)$$

We also say, for given $T \geq 0$, that a vertex v is η_T -bad if some ancestor of the vertex v at height greater than T is η -bad. Then we have the following estimate for the proportion of η_T -bad vertices:

Proposition 6.15. Fix $\varepsilon, \eta > 0$ and write $N_t^{\eta_T}$ for the collection of η_T -bad vertices at time t . Then we have

$$\mathbb{P}_x \left(\frac{N_t^{\eta_T}}{N_t} > \varepsilon \mid N_t > 0 \right) \rightarrow 0 \quad (6.13)$$

as $T \rightarrow \infty$, uniformly in $t \geq T$, for any $x \in D$.

By this we mean that for any $x \in D$, given any $\delta > 0$, there exists T' large enough that $\mathbb{P}_x(N_t^{\eta_T}/N_t > \varepsilon \mid N_t > 0) \leq \delta$ for all $t \geq T \geq T'$.

Proof. We will first show that for any $\varepsilon > 0$

$$\mathbb{P}_x(E_{T,t}^\varepsilon \mid N_t > 0) := \mathbb{P}_x \left(\frac{\sum_{i=1}^{N_t} \varphi(X_t^i) \mathbb{1}_{\{X_t^i \text{ } \eta_T\text{-bad}\}}}{\sum_{i=1}^{N_t} \varphi(X_t^i)} > \varepsilon \mid N_t > 0 \right) \rightarrow 0$$

as $T \rightarrow \infty$, uniformly in t . To do this, we will use the spine decomposition given by Lemma 4.1. Recalling the definition of \mathbb{Q}_x from this section we see that the above probability is equal to

$$\mathbb{Q}_x \left[\frac{\varphi(x)/\mathbb{P}_x(N_t > 0)}{\sum_{i=1}^{N_t} \varphi(X_t^i)} \mathbb{1}_{E_{T,t}^\varepsilon} \right] := \mathbb{Q}_x \left[Y_t \mathbb{1}_{E_{T,t}^\varepsilon} \right].$$

To see that this converges to 0 it is enough to prove that:

- $\mathbb{Q}_x(E_{T,t}^\varepsilon) \rightarrow 0$ as $T \rightarrow \infty$, uniformly in $t \geq T$ and
- For every $\delta > 0$, there exists T' and K positive, such that $\mathbb{Q}(Y_t \mathbb{1}_{|Y_t| > K}) \leq \delta$ for all $t \geq T'$.

The first point comes from the fact that the *spine particle* under this new law is unlikely to be η_T -bad for large T . More precisely, recall that $\overline{\mathbb{Q}}_x$ is a law on branching processes equipped with a distinguished path, the *spine*, such that

$$\mathbb{Q}_x = \overline{\mathbb{Q}}_x|_{\mathcal{F}_t}$$

for \mathcal{F}_t the filtration generated by the process but not the distinguished path. Under $\overline{\mathbb{Q}}_x$, we know that the spine particles evolves as a Brownian motion conditioned to remain in the domain for all time, and branches at constant rate 2λ , where each branch is either to the left or right of the spine (i.e. comes before or after in the depth-first ordering) with equal probability. Due to the mixing of this Brownian motion to a stationary distribution with density φ^2 , see Lemma 4.3, ergodicity tells us that the probability of the spine vertex being η_T -bad at time t converges to 0 as $T \rightarrow \infty$, uniformly in $t \geq T$.

The connection between the motion of the spine and the event $E_{T,t}^\varepsilon$ comes from the fact that under $\overline{\mathbb{Q}}_x$, conditioned on \mathcal{F}_t , the spine particle is chosen proportionally to φ , see (4.3). Indeed, since

$$\sum_{i=1}^{N_t} \varphi(X_t^i) \mathbb{1}_{\{X_t^i \text{ } \eta_T\text{-bad}\}} / \sum_{i=1}^{N_t} \varphi(X_t^i)$$

is positive, it is enough, for the first point, to show that its \mathbb{Q}_x expectation converges to 0 (uniformly in $t \geq T$ as $T \rightarrow \infty$). However, this follows directly from the above and (4.3) since

$$\frac{\sum_{i=1}^{N_t} \varphi(X_t^i) \mathbb{1}_{\{X_t^i \text{ } \eta_T\text{-bad}\}}}{\sum_{i=1}^{N_t} \varphi(X_t^i)} = \overline{\mathbb{Q}}_x(\text{spine } \eta_T\text{-bad at time } t | \mathcal{F}_t).$$

The second point essentially says that $(Y_t)_{t \geq 0}$ is \mathbb{Q}_x uniformly integrable. To prove it, one can use the change of measure between \mathbb{Q}_x and \mathbb{P}_x again to write

$$\mathbb{Q}_x[Y_t \mathbb{1}_{\{|Y_t| > K\}}] = \frac{\mathbb{P}_x(\{|Y_t| > K\} \cap \{N_t > 0\})}{\mathbb{P}_x(N_t > 0)} = \mathbb{P}_x(|Y_t| > K | N_t > 0).$$

Since $\varphi(x)/t\mathbb{P}_x(N_t > 0)$ is uniformly bounded above for $t \geq 1$ say, by Theorem 1.4, we just need to show that for any $\delta > 0$ there exists K and T' such that

$$\sup_{t \geq T'} \mathbb{P}_x \left(\frac{\sum_{i=1}^{N_t} \varphi(X_t^i)}{t} < 1/K \mid N_t > 0 \right) \leq \delta. \quad (6.14)$$

However, this is a direct consequence of the convergence given by Theorem 1.5, since we know that for fixed K the probability above converges, as $t \rightarrow \infty$, to the probability that an exponential random variable is less than $1/K$.

We must now deduce that

$$\mathbb{P}_x \left(\frac{N_t^{\eta_T}}{N_t} > \varepsilon \mid N_t > 0 \right) \rightarrow 0$$

uniformly in $t \geq T$ as $T \rightarrow \infty$ from the fact that

$$\mathbb{P}_x(E_{T,t}^\varepsilon | N_t > 0) \rightarrow 0.$$

The idea behind this is that $\sum_{i=1}^{N_t} \varphi(X_t^i) \mathbb{1}_{\{X_t^i \text{ } \eta_T\text{-bad}\}} / \sum_{i=1}^{N_t} \varphi(X_t^i)$ is a reasonable approximation to $N_t^{\eta_T}/N_t$ on survival at large times. By Corollary 1.7, we know that for any $\delta > 0$ there exist r and T' positive such that

$$\sup_{t \geq T'} \mathbb{P}_x \left(\frac{N_t^{Dr}}{N_t} > \frac{\varepsilon}{2} \mid N_t > 0 \right) \leq \delta/2 \quad (6.15)$$

where $D_r = \{y \in D : \varphi(y) < 1/r\}$ and $N_t^{D_r}$ is the number of particles in D_r at time t . Also, write $N_t^{D_r^c, \eta_T}$ for the number of particles that are η_T -bad and lie in D_r^c at time t . Then

$$\frac{\sum_{i=1}^{N_t} \varphi(X_t^i)}{N_t} \leq \|\varphi\|_\infty \quad \text{and} \quad \frac{N_t^{D_r^c, \eta_T}}{\sum_{i=1}^{N_t} \varphi(X_t^i) \mathbb{1}_{\{X_t^i \eta_T\text{-bad}\}}} \leq r.$$

Bounding $N_t^{\eta_T}$ above by $N_t^{D_r} + N_t^{D_r^c, \eta_T}$, and choosing $T \geq T'$ large enough that $\mathbb{P}_x(E_{T,t}^{(2\|\varphi\|_\infty r)^{-1}\varepsilon}) \leq \delta/2$, we have that

$$\mathbb{P}_x \left(\frac{N_t^{\eta_T}}{N_t} > \varepsilon \mid N_t > 0 \right) \leq \delta$$

for all $t \geq T$. □

Remark 6.16. We can extend the proof to show that for any $c > 0$ and $x \in D$,

$$\mathbb{P}_x \left(\left\{ \frac{N_{ct}^{\eta_T}}{N_{ct}} > \varepsilon \right\} \cap \{N_{ct} > 0\} \mid N_t > 0 \right) \rightarrow 0 \quad (6.16)$$

uniformly in $ct \wedge t \geq T$ as $T \rightarrow \infty$. This follows immediately from (6.13), since

$$\mathbb{P}_x \left(\left\{ \frac{N_{ct}^{\eta_T}}{N_{ct}} > \varepsilon \right\} \cap \{N_{ct} > 0\} \mid N_t > 0 \right) = \mathbb{P}_x \left(\left\{ \frac{N_{ct}^{\eta_T}}{N_{ct}} > \varepsilon \right\} \cap \{N_{ct} > 0\} \mid N_{ct} > 0 \right) \frac{\mathbb{P}_x(N_{ct} > 0)}{\mathbb{P}_x(N_t > 0)}$$

where $\mathbb{P}_x(N_{ct} > 0)/\mathbb{P}_x(N_t > 0)$ is uniformly bounded in $ct \wedge t \geq T$.

Remark 6.17. For the more general set up, the above arguments do not need to be changed, except to replace $\lambda \int_D \varphi(y)^3 dy$ in (6.12) by

$$\frac{\lambda \mathbb{E}[A^2 - A]}{2(m-1)} \int_D \varphi(y)^3 dy$$

where A is the offspring distribution and φ as usual becomes the first eigenfunction of the generator. This follows from Remark 4.2, since branching occurs along the spine at rate $(m/m-1)\lambda$ and the number of younger siblings at each such point has expectation $\mathbb{E}[(A^2 - A)/2]$.

The next Lemma provides the key connection between \hat{S} and the height function, for a branching Brownian motion tree that is conditioned to survive for a long time.

Lemma 6.18. Write $\bar{\mathbb{P}}_x$ for the law of a branching Brownian motion started at x , plus a vertex u chosen uniformly from it. Then for any $\eta > 0$ we have

$$\bar{\mathbb{P}}_x(u \text{ is } \eta_T\text{-bad} \mid N_t > 0) = 0,$$

uniformly in $t \geq T$ as $T \rightarrow \infty$.

Proof. By conditioning on N_s and $N_s^{\eta_T}$ for all $0 \leq s < \infty$ and L , which is the total length of the tree as usual, we see that

$$\bar{\mathbb{P}}_x(u \text{ is } \eta_T\text{-bad} \mid N_t > 0) = \mathbb{E}_x \left[\int_0^\infty \frac{N_s^{\eta_T}}{N_s} \frac{N_s}{L} ds \mid N_t > 0 \right].$$

Thus we need to show that, given $\varepsilon > 0$, we have

$$\mathbb{E}_x \left[\int_0^\infty \mathbb{1}_{\{|N_s^{\eta_T}/N_s| > \varepsilon/2\}} \frac{N_s}{L} ds \mid N_t > 0 \right] \leq \varepsilon/2$$

for all $t \geq T$, whenever T is large enough. First note that by (6.11), we can pick a c such that $\mathbb{P}_x(L \leq ct^2 \mid N_t > 0) \leq \varepsilon/8$ for all t , and also using our asymptotic for the survival probability, can choose an R such that $\bar{\mathbb{P}}_x(N_{Rt} > 0 \mid N_t > 0) \leq \varepsilon/8$ for all t . Moreover, since

$$\mathbb{E}_x \left[\int_0^s N_s ds \mid N_t > 0 \right] \leq Mst$$

for some M and all s, t , we have that $\mathbb{E}_x[\int_0^{bt} N_s ds | N_t > 0] \leq Mbt^2$ and can therefore choose $b > 0$ small enough that this is less than $\varepsilon ct^2/8$. Combining this with the condition on L , our problem is reduced to showing that

$$\lim_{T \rightarrow \infty} \frac{1}{ct^2} \mathbb{E}_x \left[\int_{bt}^{Rt} \mathbb{1}_{\{|N_s^{\eta_T}/N_s| > \varepsilon/2\}} N_s ds \mid N_t > 0 \right] \leq \varepsilon/8$$

uniformly in $t \geq T$. However, by Remark 6.16 we know that

$$\sup_{bt \geq T} \sup_{s \in [bt, Rt]} \mathbb{P}_x(\{|N_s^{\eta_T}/N_s| > \varepsilon/2\} \cap \{N_s > 0\} \mid N_t > 0) \rightarrow 0$$

as $T \rightarrow \infty$ (the uniformity comes straight from the proof.) Hence, applying Fubini, Cauchy-Schwarz and integrating, similarly to in the proof of Proposition 6.8, we obtain the result. \square

We will use the above to show that, if we choose k particles (u_1, \dots, u_k) uniformly from a branching Brownian motion tree conditioned to survive up to time t and define the matrices

$$(d_t^S(u_i, u_j))_{1 \leq i < j \leq k} = \frac{\hat{S}(u_i) + \hat{S}(u_j) - 2\hat{S}(v_{i,j})}{t}$$

and

$$(d_t^H(u_i, u_j))_{1 \leq i < j \leq k} = \frac{h(u_i) + h(u_j) - 2h(v_{i,j})}{t}$$

where $v_{i,j}$ is the most recent common ancestor of u_i and u_j , then the two are essentially the same up to the multiplicative constant $\lambda \int_D \varphi(y)^3 dy$. Note that d_t^H is actually how we define distances in the genealogical tree (after rescaling by t) and

$$\lambda \int_D \varphi(y)^3 dy = \frac{\sigma}{\alpha}$$

where α is the constant from the statement of Theorem 1.1.

Proposition 6.19. *Let d_t^S , d_t^H and (u_1, \dots, u_k) be defined as above. Then for any $\varepsilon > 0$*

$$\mathbb{P} \left(\left\| \left(\frac{\sigma}{\alpha} d_t^H(u_i, u_j) - d_t^S(u_i, u_j) \right) \right\| > \varepsilon \mid N_t > 0 \right) \rightarrow 0 \quad (6.17)$$

as $t \rightarrow \infty$, where the distance is the Euclidean distance between $k \times k$ matrices.

Proof. We prove this in the case $k = 2$, the general result following by a union bound. Lemma 6.18 tells us that if we pick a vertex u uniformly at random from a tree conditioned to survive to time t , then with high probability the value of $\hat{S}(u)/h(u)$ is close to σ/α . In fact, the probability that this holds at *all* points along u 's ancestry (except for at very small times) is very large. Hence, if we pick any pair of vertices uniformly, since d_t^H and d_t^S depend only on their joint ancestry and the values of \hat{S} along it, they will be close (when d_t^H is multiplied σ/α) with high probability.

To make this more precise, pick an $\eta > 0$. Then we can pick T large enough that the probability of two uniformly chosen vertices u_1, u_2 from the tree under B_n^H being η_T -bad is arbitrarily small, uniformly in $t \geq T$. We can also pick R large enough the probability that $[h(z_1) \wedge h(z_2), h(z_1) \vee h(z_2)] \subset [t/R, Rt]$ is arbitrarily close to 1, uniformly in t (see the proof of Lemma 6.18.)

Then, on the event that u_1 and u_2 are both η_T -good, and $[h(u_1) \wedge h(u_2), h(u_1) \vee h(u_2)] \subset [t/R, Rt]$, we have, as long as $t \geq RT$

$$\left| \frac{\hat{S}(u_i)}{h(u_i)} - \sigma/\alpha \right| \leq \eta$$

for $i = 1, 2$. The other contributions to d_t^H and d_t^S come from the height (resp. the value of \hat{S}) at the most recent common ancestor u of the two vertices. There are two possibilities: either $h(u) \leq T$, or not. In the second case we have that

$$\left| \frac{\hat{S}_n(u)}{h(u)} - \sigma/\alpha \right| \leq \eta$$

and so

$$\left| d_t^S(u_1, u_2) - \frac{\sigma}{\alpha} d_t^H(u_1, u_2) \right| \leq 4R\eta.$$

In the first we have

$$\left| d_t^S(u_1, u_2) - \frac{\sigma}{\alpha} d_t^H(u_1, u_2) \right| \leq 2R\eta + \left(\frac{2\sigma}{\alpha} + \sup_{s \leq T} \frac{N_s}{s} \right) \frac{T}{t}.$$

However for fixed T , we have that $\mathbb{P}_x \left(\sup_{s \leq T} |N_s/s| \geq K \mid N_t > 0 \right)$ converges to 0 as $K \rightarrow \infty$, uniformly in t (see Proposition 1.5). This proves the convergence in probability. \square

Immediately from the proof, we also get the following Corollary, which we will need later on.

Corollary 6.20. *We also have for any $\varepsilon > 0$ and $c > 0$*

$$\mathbb{P} \left(\left\| \left(\frac{\sigma}{\alpha} d_{ct}^H(u_i, u_j) - d_{ct}^S(u_i, u_j) \right) \right\| > \varepsilon \mid N_t > 0 \right) \rightarrow 0 \quad (6.18)$$

as $t \rightarrow \infty$, where the distance is the Euclidean distance between $k \times k$ matrices.

Remark 6.21. *The proofs of Lemma 6.18 and Proposition 6.19, given all the previous work, are exactly the same for the more general set up.*

6.3 Convergence to the CRT

6.3.1 Preliminaries on converge of metric measure spaces

Before we can prove Theorem 1.1, we must introduce various notions of convergence for metric spaces, and more generally, for metric measure spaces. Although our aim is to prove convergence of conditioned genealogical trees in the sense of Gromov-Hausdorff distance between metric spaces, it turns out to be helpful to go through the framework of metric measure spaces. We first recall the definition of the Gromov-Hausdorff metric on \mathbb{X}_c : the space of (isometry classes of) compact metric spaces.

Definition 6.22. *The Gromov-Hausdorff distance between (X, r_X) and (Y, r_Y) in \mathbb{X}_c is given by*

$$d_{GH}((X, r_X), (Y, r_Y)) = \inf_{g_X, g_Y, Z} d_H^{(Z, r_Z)}(g_X(X), g_Y(Y)),$$

where the infimum is taken over all isometric embeddings g_X, g_Y from X and Y to a common metric space (Z, r_Z) , and $d_H^{(Z, r_Z)}$ is the usual Hausdorff distance on (Z, r_Z) .

Now we will briefly discuss some modes of convergence for metric measure spaces, and how they are related, both with each other and the above. For us, a *metric measure space* (X, r, μ) will be a compact metric space, equipped with a finite Borel measure. These will be considered up to isometry, where $(X, r, \mu) \sim (X', r', \mu')$ if there exists a measure preserving isometry between X and X' . We denote the set of (isometry classes) of these spaces by \mathbb{X} . We will be interested in the *Gromov-Prohorov* metric and the *Gromov-Hausdorff-Prohorov* metric on \mathbb{X} . We will begin by defining the so-called *Gromov-Weak* topology.

Definition 6.23. [GPW09, Definition 2.3] *We will call a function $\Phi : \mathbb{X} \rightarrow \mathbb{R}$ a polynomial if there exists an $k \in \mathbb{N}$ and a bounded continuous function $\phi : [0, \infty)^{\binom{k}{2}} \rightarrow \mathbb{R}$ such that*

$$\Phi((X, r, \mu)) = \int \mu^{\otimes k}(d(x_1, \dots, x_n)) \phi((r(x_i, x_j))_{1 \leq i < j \leq k}),$$

where $\mu^{\otimes k}$ is the product measure of μ . Write Π for the set of all polynomials.

Definition 6.24. [GPW09, Definition 2.8] *A sequence $\mathcal{X}_n \in \mathbb{X}$ is said to converge to $\mathcal{X} \in \mathbb{X}$ with respect to the Gromov-weak topology if and only if $\Phi(\mathcal{X}_n)$ converges to $\Phi(\mathcal{X})$ in \mathbb{R} , for all polynomials $\Phi \in \Pi$.*

It was proved in [GPW09, Theorem 5] that this topology is metrised by the *Gromov-Prohorov metric* defined below.

Definition 6.25. *The Gromov-Prohorov distance between $\mathcal{X} = (X, r_X, \mu_X)$ and $\mathcal{Y} = (Y, r_Y, \mu_Y)$ in \mathbb{X} is given by*

$$d_{GP}(\mathcal{X}, \mathcal{Y}) = \inf_{g_X, g_Y, Z} d_{Pr}^{(Z, r_Z)}((g_X)_*(\mu_X), (g_Y)_*(\mu_Y)),$$

where the infimum is as in Definition 6.22 and $d_P^{(Z, r_Z)}$ is the Prohorov distance between probability measures on (Z, r_Z) .

Finally, we define the Gromov-Hausdorff-Prohorov metric [ADH13],[Mie09] on \mathbb{X} .

Definition 6.26. *Let \mathcal{X}, \mathcal{Y} be as in Definition 6.25. The Gromov-Hausdorff-Prohorov distance between \mathcal{X} and \mathcal{Y} is defined by*

$$d_{GHP}(\mathcal{X}, \mathcal{Y}) = \inf_{g_X, g_Y, Z} \left(d_{Pr}^{(Z, r_Z)}((g_X)_*(\mu_X), (g_Y)_*(\mu_Y)) + d_H^{(Z, r_Z)}(g_X(X), g_Y(Y)) \right).$$

Remark 6.27. *It is clear from the above definitions that convergence in the Gromov-Hausdorff-Prohorov metric implies convergence in both the Gromov-Hausdorff metric and the Gromov-Prohorov metric.*

We will need a couple of facts for our proof:

Lemma 6.28. [GPW09, Corollary 3.1] *A sequence $\{\mathbb{P}_n\}_{n \in \mathbb{N}}$ of probability measures on \mathbb{X} converges weakly to a probability measure \mathbb{P} with respect to the Gromov-weak topology, if and only if*

- (i) *The family $\{\mathbb{P}_n\}_{n \in \mathbb{N}}$ is relatively compact in the space of probability measures on \mathbb{X} .*
- (ii) *For all polynomials $\Phi \in \Pi$, $\mathbb{P}_n[\Phi] \rightarrow \mathbb{P}[\Phi]$ in \mathbb{R} as $n \rightarrow \infty$.*

and

Lemma 6.29. [ADH13, Theorem 2.4], [BBI01, Theorem 7.4.15] *A set $\mathbb{K} \subset \mathbb{X}$ is relatively compact with respect to the Gromov-Hausdorff-Prohorov metric if and only if*

- (i) *There is a constant D such that $\text{diam}(\mathcal{X}) < D$ for all $\mathcal{X} \in \mathbb{K}$.*
- (ii) *For every $\delta > 0$ there exists $N = N_\delta$ such that for all $\mathcal{X} \in \mathbb{K}$, X can be covered by N_δ balls of radius δ .*
- (iii) $\sup_{\mathcal{X} \in \mathbb{K}} \mu_X(X) < +\infty$

6.3.2 Proof of the main theorem

In this section we will assume that our critical branching Brownian motion is always started from position $x \in D$. Recall that for any $y > 0$ we let \mathcal{T}_n^y be the genealogical tree generated by this process, when it is conditioned to survive until time ny . This is just our usual tree, but with the marks forgotten. Denote by d_n^y the natural metric on this tree (given by the length of the path connecting two vertices by their most recent common ancestor). Finally, write $(\mathcal{T}_{e^y}, d_{e^y})$ for the real tree whose contour function is given by e^y , a Brownian excursion conditioned to reach height y . We will often want to put a *uniform measure* on these trees, to make them into metric measure spaces. In the case of \mathcal{T}_n^y the measure will be denoted by μ_n^y and will be defined by the following procedure. Let $\phi : \mathcal{T}_n^y \rightarrow \mathbb{R}$ send the vertex visited at time t in a depth-first exploration of the tree to $t \in \mathbb{R}$. Then we can put a uniform measure on the image of ϕ (which will almost surely be a finite interval) and pull this back to the tree. Note that choosing a vertex from \mathcal{T}_n^y according to μ_n^y is the same as choosing a vertex uniformly from the tree, as in Section 6.2. To define the uniform measure μ_{e^y} for $(\mathcal{T}_{e^y}, d_{e^y})$, recall that \mathcal{T}_{e^y} is the real tree encoded by the excursion e^y . This means that the metric space is (isometric to) the interval on which the excursion is supported, quotiented by the equivalence

relation that $s \sim r$ if $\inf\{e_u^y; u \in [r, s]\} = e_r^y = e_s^y$, with metric $d_{e^y}(r, s) = e_r^y + e_s^y - 2 \inf\{e_u^y; u \in [r, s]\}$. The uniform measure μ_{e^y} is then just the quotient measure of uniform measure on the interval.

Recall that, since we are working in the case of binary branching Brownian motion, we would like to show that for any $y > 0$

$$(\mathcal{T}_n^{\alpha y}, \frac{1}{\alpha n} d_n^{\alpha y}) \xrightarrow{n \rightarrow \infty} (\mathcal{T}_{e^y}, d_{e^y})$$

in distribution, with respect to the Gromov-Hausdorff distance, where

$$\alpha = \sqrt{\frac{2}{\lambda(1, \varphi) \int_D \varphi(y)^3 dy}}.$$

Proof. An outline of the proof is as follows:

- (1) We will show that if we make $(\mathcal{T}_n^{\alpha y}, \frac{1}{\alpha n} d_n^{\alpha y})$ into a metric measure space for each n , by equipping it with the uniform measure $\mu_n^{\alpha y}$, then the family

$$\left(\mathcal{T}_n^{\alpha y}, \frac{1}{\alpha n} d_n^{\alpha y}, \mu_n^{\alpha y} \right)_{n \geq 0}$$

is tight with respect to the Gromov-Hausdorff-Prohorov metric.

- (2) Using Propositions 6.10 and 6.19, we will show convergence of the above family, with respect to the Gromov-Prohorov metric, to $(\mathcal{T}_{e^y}, d_{e^y}, \mu_{e^y})$, where μ_{e^y} is uniform measure on the real tree \mathcal{T}_{e^y} , as defined above.
- (3) This also characterises the subsequential limits with respect to the Gromov-Hausdorff-Prohorov metric, since Gromov-Hausdorff-Prohorov convergence implies Gromov-Prohorov convergence. Thus we have the convergence in distribution

$$\left(\mathcal{T}_n^{\alpha y}, \frac{1}{\alpha n} d_n^{\alpha y}, \mu_n^{\alpha y} \right) \xrightarrow{n \rightarrow \infty} (\mathcal{T}_{e^y}, d_{e^y}, \mu_y)$$

with respect to the Gromov-Hausdorff-Prohorov metric. Consequently, by Remark 6.27, we have that

$$\left(\mathcal{T}_n^{\alpha y}, \frac{1}{\alpha n} d_n^{\alpha y} \right) \xrightarrow{n \rightarrow \infty} (\mathcal{T}_{e^y}, d_{e^y})$$

with respect to the Gromov-Hausdorff metric, as desired.

All that remains is to verify the statements in (1) and (2). For Part (1), we need to show that for any $\varepsilon > 0$ there exists a relatively compact $\mathbb{K} \subset \mathbb{X}$ (wrt the Gromov-Hausdorff metric) such that

$$\inf_n \mathbb{P}((\mathcal{T}_n^{\alpha y}, \frac{1}{\alpha n} d_n^{\alpha y}) \in \mathbb{K}) \geq 1 - \varepsilon.$$

We will use the characterisation given by Lemma 6.29. Since all of our measures are probability measures, condition (iii) of this characterisation is trivial. We begin by proving the existence of $K > 0$ such that

$$\sup_n \mathbb{P} \left(\text{diam} \left(\mathcal{T}_n^{\alpha y}, \frac{1}{\alpha n} d_n^{\alpha y} \right) > 2K \right) < \varepsilon/2, \quad (6.19)$$

which gives condition (ii). However, $\text{diam}(\mathcal{T}_n^{\alpha y}, \frac{1}{\alpha n} d_n^{\alpha y})$ is less than $2/\alpha n$ times the maximum height of a branching Brownian motion process conditioned to survive until time $\alpha n y$, so this follows immediately from our asymptotic for the survival probability. To complete the proof of tightness we will consider the tree $(\mathcal{T}_n^{\alpha y}, \frac{1}{\alpha n} d_n^{\alpha y})$ cut off at height Kn , and show that there exists an $M > 0$ such that the probability of this cut off tree having a δ -net with less than M/δ^4 balls is greater than $1 - \delta\varepsilon/2$ for all $n \in \mathbb{N}$ and $\delta > 0$. By summing probabilities over the sequence $\delta_k = 2^{-k}$, and combining with (6.19) this provides the set \mathbb{K} that we need. This estimate is of course very crude, but will suffice for our purposes here.

To prove the claim, we will take K as above, and for any $\delta > 0, n \in \mathbb{N}$ divide $[0, Kn]$ into intervals of length $n\delta/2 := b_{n,\delta}$ (so there are $2K/\delta$ of them.) Then we will choose M such that in any one of these intervals $[jb_{n,\delta}, (j+1)b_{n,\delta}]$ the probability of having more than $M/2K\delta^3$ particles at time $jb_{n,\delta}$ with descendants at time $(j+1)b_{n,\delta}$ is less than $\delta^2\varepsilon/4K$, independently of j, n, δ . Assuming we can do this, summing up gives that the probability of this holding for *any* of the subintervals is less than $\delta\varepsilon/2$. Moreover, on the event that it doesn't happen for any subinterval, we can put a δ ball, for each j , at every vertex in level $jb_{n,\delta}$ that has a descendant at level $(j+1)b_{n,\delta}$. This will cover the tree up to level Kn , since a δ -ball is effectively a δn ball (remember that lengths are rescaled), and the choice of positions means that every vertex with height in $[jb_{n,\delta}, (j+1)b_{n,\delta}]$ for $j \geq 1$ is covered by one of the δ balls centred at level $(j-1)b_{n,\delta}$. For $j = 0$, vertices in this interval are covered by the ball placed at the root. In this covering, by definition of the event, there are less than M/δ^4 balls. Thus it remains to show that we can choose such an M , independently of δ, n . This is a result of the following observations:

- By Theorem 1.4, there exists an $R > 0$ such that the probability of a critical branching Brownian motion starting from z surviving until time $n\delta/2$ is less than $R/n\delta$ for all n, δ and z .
- Let $\tilde{N}_{jb_{n,\delta}}$ be the number of vertices at time $jb_{n,\delta}$ that have descendants at time $(j+1)b_{n,\delta}$. Then, by Wald's identity, we have for all j, n and δ

$$\mathbb{P}_x \left(\tilde{N}_{jb_{n,\delta}} \geq \frac{M}{2K\delta^3} \right) \leq \frac{2RK\delta^2 \mathbb{E}_x[N_{jb_{n,\delta}}]}{Mn} \leq \frac{2CRK\delta^2}{Mn}$$

where $C := \sup_{z,u} \mathbb{E}_z[N_u]$ is finite by the proof of Theorem 1.6. Note that we are not conditioning on survival here.

Therefore,

$$\mathbb{P}_x \left(\tilde{N}_{jb_{n,\delta}} \geq \frac{M}{2K\delta^3} \middle| N_{\alpha yn} > 0 \right) \leq \frac{2CRK\delta^2}{Mn} \times \mathbb{P}_x(N_{\alpha yn} > 0)^{-1}.$$

This is less than $\delta^2\varepsilon/4K$ if we choose

$$M = \frac{8CRK^2}{\varepsilon} \sup_{n \geq 1} (n\mathbb{P}_x(N_{\alpha yn} > 0))^{-1}$$

where the supremum is finite by Theorem 1.4.

Now we move on to (2). We would like to show that $(\mathcal{T}_n^{\alpha y}, \frac{1}{\alpha n} d_n^{\alpha y}, \mu_n^{\alpha y})$ converges to $(\mathcal{T}_{e^y}, d_{e^y}, \mu_y)$ in the Gromov-Prohorov metric, or equivalently, with respect to the Gromov-weak topology. We will consider the latter formulation, in order to use the characterisation given by Lemma 6.28. Since convergence in the Gromov-Hausdorff-Prohorov sense implies convergence in the Gromov-Prohorov/Gromov-Weak sense, the argument for (1) also shows that part (i) of the characterisation (relative compactness of the laws) is satisfied. Therefore, we need only show for any polynomial $\Phi \in \Pi$, writing \mathbb{P}_n^H for the law of $(\mathcal{T}_n^{\alpha y}, \frac{1}{\alpha n} d_n^{\alpha y}, \mu_n^{\alpha y})$ and \mathbb{P} for the law of $(\mathcal{T}_{e^y}, d_{e^y}, \mu_y)$, that we have $\mathbb{P}_n^H[\Phi] \rightarrow \mathbb{P}[\Phi]$ as $n \rightarrow \infty$. To do this, we use the convergence of the martingale, Proposition 6.10, and Proposition 6.19.

Fix a polynomial Φ from $[0, \infty)^{\binom{k}{2}}$ to \mathbb{R} . We will first introduce some notation. We let B_n^H be the event that a branching Brownian motion tree reaches height αyn , and B_n^S be the event that its associated process \hat{S} , recall the definition from Definition 6.3, reaches level σyn . Note that $\sigma/\alpha = \lambda \int_D \varphi(y)^3 dy$, which is the constant from Propositions 6.15 and 6.19 that relates the height function and \hat{S} . We denote by \mathbb{P}_n^S the law (on metric measure spaces) which is the same as \mathbb{P}_n^H , except that the tree is conditioned on B_n^S rather than B_n^H . To prove that $\mathbb{P}_n^H[\Phi] \rightarrow \mathbb{P}[\Phi]$ we show that:

- (i) $\mathbb{P}_x(B_n^S | B_n^H) \rightarrow 1$ and $\mathbb{P}_x(B_n^H | B_n^S) \rightarrow 1$ as $n \rightarrow \infty$.
- (ii) $\mathbb{P}_n^S[\Phi] \rightarrow \mathbb{P}[\Phi]$ as $n \rightarrow \infty$.

Combining these provides the result. We start with (i). The convergence $\mathbb{P}_x(B_n^S | B_n^H) \rightarrow 1$ is in fact a direct consequence of Proposition 6.15, and the fact that given a branching Brownian motion survives until time $y\alpha n$, the probability of it surviving to time $\varepsilon y\alpha n$ and having more than two particles alive at this time, is high. For the second part, it is therefore sufficient to show that $\mathbb{P}_x(B_n^H)/\mathbb{P}_x(B_n^S) \rightarrow 1$ as $n \rightarrow \infty$. However, the convergence given by Proposition 6.10 allows us to compute an exact asymptotic for $\mathbb{P}_x(B_n^S)$, just as in for example [Gal05, Section 1.4, p263]. It is easy to check that this is indeed equal to $\frac{\varphi(z)}{y\sigma n} = \frac{\varphi(x)}{\alpha y n \int_D \varphi(z)^3 dz}$.

For the proof of (ii) recall that we write \hat{S}^y for the process \hat{S} conditioned to reach y . This is where we will use Proposition 6.10. This, along with the Skorokhod representation theorem, tells us that there exists a sequence of processes $(Z_t^{(n)})_{t \geq 0}$, equal in distribution to $(\hat{S}_{n^2 t}^{y\sigma n}/n)_{t \geq 0}$, such that

$$(Z_t^{(n)})_{t \geq 0} \rightarrow (\sigma e_t^y)_{t \geq 0} \quad (6.20)$$

uniformly almost surely as $n \rightarrow \infty$. Here we set the processes identically equal to zero after they reach zero, at times that we denote by $(\tau_n)_{n \in \mathbb{N}}$ and τ . Choose k points $(z_i; 1 \leq i \leq k)$ uniformly from $[0, \tau]$ and for each n set $(z_i^n; 1 \leq i \leq k) = \frac{\tau_n}{\tau}(z_i; 1 \leq i \leq k)$. Then for each n , the law of

$$\left((Z_t^{(n)})_{t \geq 0}, (z_i^n; 1 \leq i \leq k) \right)$$

is that of $(\hat{S}_{n^2 t}^{y\sigma n}/n)_{t \geq 0}$ together with k points chosen uniformly from its length. Define the distance between z_i^n and z_j^n for $1 \leq i < j \leq k$ by

$$d_n^Z(z_i^n, z_j^n) := Z_{z_i^n}^{(n)} + Z_{z_j^n}^{(n)} - 2 \inf_{s \in [z_i^n, z_j^n]} Z_s^{(n)}. \quad (6.21)$$

Setting

$$d_{ey}(z_i, z_j) = e_{z_i}^y + e_{z_j}^y - 2 \inf_{s \in [z_i, z_j]} e_s^y$$

(corresponding to the normal metric d_{ey} in \mathcal{T}_{ey}) it is immediate from (6.20) that

$$(z_i^n; 1 \leq i \leq k) \longrightarrow (z_i; 1 \leq i \leq k)$$

almost surely as $n \rightarrow \infty$, and so also that

$$\frac{1}{\sigma} (d_n^Z(z_i^n, z_j^n))_{1 \leq i < j \leq k} \longrightarrow (d_{ey}(z_i, z_j))_{1 \leq i < j \leq k} \quad (6.22)$$

almost surely as $n \rightarrow \infty$. This is useful, because the law of the object on the right is the same as the law of the matrix of pairwise distances between k points chosen independently according to μ_{ey} from $(\mathcal{T}_{ey}, d_{ey})$. Note that although the metric space is actually a quotient of $[0, \tau]$ in this case, the law above is not affected.

Moreover, the law of the object on the left is what we get if we choose k vertices uniformly from a branching Brownian motion tree conditioned on B_n^S , and take the matrix defined by d_n^S in Proposition 6.19. However, we know by Proposition 6.19 and Corollary 6.20 that if we condition instead on B_n^H , the difference between this matrix and the matrix of pairwise distances under $\frac{1}{\alpha n} d_n^{\alpha y}$, converges to 0 in probability as $n \rightarrow \infty$. By part (i), this is still true under conditioning on B_n^S . Since Φ is continuous and bounded, the proof of part (ii) is complete. \square

We conclude by noticing that, since the above proof does not use anything specific about the branching diffusion (except the results that have been proven earlier in the paper and some constants), the more general statement of Theorem 1.1 holds.

Remark 6.30. *The same result holds for a branching diffusion with generator L and offspring distribution A as in Theorem 1.1 where*

$$\alpha = \sqrt{\frac{4(m-1)}{\lambda(1, \varphi) \mathbb{E}[A^2 - A] \int_D \varphi(y)^3 dy}}$$

so that

$$\frac{\sigma}{\alpha} = \frac{\lambda \mathbb{E}[A^2 - A]}{2(m-1)} \int_D \varphi(y)^3 dy$$

as in Remark 6.17.

References

- [ADH13] R. Abraham, J.F. Delmas, and P. Hoscheit. A note on the Gromov-Hausdorff-Prokhorov distance between (locally) compact metric measure spaces. *Electronic Journal of Probability*, 18, 2013.
- [AH83] S. Asmussen and H. Hering. *Branching processes*. Birkhäuser, Boston, MA, 1983.
- [Ald91] D. Aldous. The continuum random tree I. *Annals of Probability*, 19:1–28, 1991.
- [Ald93] D. Aldous. The continuum random tree III. *Annals of Probability*, 21:248–289, 1993.
- [Bañ99] R. Bañuelos. Intrinsic ultracontractivity and eigenfunction estimates for Schrödinger operators. *Journal of Functional Analysis*, 100:181–205, 1999.
- [BBI01] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*. AMS, Boston MA, 2001.
- [BBS11] J. Berestycki, N. Berestycki, and J. Schweinsberg. Survival of near-critical branching Brownian motion with absorption. *Journal of Statistical Physics*, 143:833–854, 2011.
- [BBS14] J. Berestycki, N. Berestycki, and J. Schweinsberg. Critical branching Brownian motion with absorption: survival probability. *Probability Theory and Related Fields*, 160:489–520, 2014.
- [Bil95] P. Billingsley. *Probability and Measure*. John Wiley and Sons, New York, 3rd edition, 1995.
- [CR88] B. Chauvin and A. Rouault. KPP equation and supercritical branching Brownian motion in the subcritical speed area. Application to spatial trees. *Probability Theory and Related Fields*, 80:299–314, 1988.
- [DN80] D. J. Daley and P. Narayan. *Series Expansions of Probability Generating Functions and Bounds for the Extinction Probability of a Branching Process*, volume 17, no.4. 1980.
- [Doo57] J.L. Doob. Conditioned Brownian motion and the boundary limits of harmonic functions. *Bulletin of the Mathematical Society of France*, 85:431–458, 1957.
- [DS84] E.B. Davies and B. Simon. Ultracontractivity and heat kernels for Schrödinger operators and Dirichlet Laplacians. *Journal of Functional Analysis*, 59:335–395, 1984.
- [EK04] J. Engländer and A.E. Kyprianou. Local extinction vs. local exponential growth for spatial branching processes. *Annals of Probability*, 32:78–99, 2004.
- [Eva98] L.C. Evans. *Partial Differential Equations*. Graduate Studies in Mathematics, Volume 19. AMS, 1998.
- [Gal05] J.F. Le Gall. Random trees and applications. *Probability Surveys*, 2, 2005.
- [GD02] J.F. Le Gall and T. Duquesne. Random trees, Lévy processes and spatial branching processes. *Astérisque*, 281, 2002.
- [GPW09] A. Greven, P. Pfaffelhuber, and A. Winter. Convergence in distribution of random metric measure spaces (λ -coalescent measure trees). *Probability Theory and Related Fields*, 145(1-2):285–322, 2009.

- [GT83] D. Gilbarg and N.S. Trudinger. *Elliptic partial differential equations of second order*. Springer-Verlag, Berlin, 2nd edition, 1983.
- [HH07] J.W. Harris and S.C. Harris. Survival probabilities for branching Brownian motion with absorption. *Electronic Communications in Probability*, 12:81–92, 2007.
- [HH09] R. Hardy and S.C. Harris. A spine approach to branching diffusions with applications to L_p -convergence of martingales, pages 280–330. Séminaire de Probabilité XLII, volume 1979, Lecture Notes in Math. Springer, Berlin, 2009.
- [HHK12] S.C. Harris, M. Hesse, and A.E. Kyprianou. Branching Brownian motion in a strip: Survival near criticality. *Annals of Probability*, 44 (1):235–275, 2012.
- [HR14] S.C. Harris and M. Roberts. A strong law of large numbers for branching processes: almost sure spine events. *Electronic Communications in Probability*, 19,28, 2014.
- [HR15] S.C. Harris and M. Roberts. The many-to-few lemma and multiple spines. *Annales de l’Institut Henri Poincaré (B) Probability and Statistics*, 2015.
- [JS87] J.Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, 1987.
- [Kes78] H. Kesten. Branching Brownian motion with absorption. *Stochastic Processes and Applications*, 7:9–47, 1978.
- [Kol38] A.N. Kolmogorov. Zur Lösung einer biologischen Aufgabe. *Izv. Naučn.-Issled. Inst. Mat. i Meh. Tomsk. Gosudarstv. Univ.*, 2:1–6, 1938.
- [McK75] H.P. McKean. Application of Brownian motion to the equation of Kolmogorov-Petrovski-Piscounov. *Communications in Pure and Applied Mathematics*, 28:323–331, 1975.
- [Mie08] G. Miermont. Invariance principles for spatial multitype Galton-Watson trees. *Annales de l’institut Henri Poincaré (B) Probabilités et Statistiques*, 44.6:1128–1161, 2008.
- [Mie09] G. Miermont. Tessellations of random maps of arbitrary genus. *Ann. Sci. Éc. Norm. Supér.*, 42(5:725–781, 2009.
- [MNV09] M. Meerschaert, E. Nane, and P. Vellaisamy. Fractional cauchy problems on bounded domains. *Annals of Probability*, 37:997–1007, 2009.
- [Pin85] R.G. Pinsky. On the convergence of diffusion processes conditioned to remain in a bounded region for large time to limiting positive recurrent diffusion processes. *Annals of Probability*, 13:363–378, 1985.
- [Rob10] M.I. Roberts. Spine changes of measure and branching diffusions. *PhD thesis, University of Bath*, <http://people.bath.ac.uk/mir20/thesis.pdf>, 2010.
- [Rud76] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill Book Co., New York, 3rd edition, 1976.
- [RY91] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer, 1991.
- [Sev58] B.A. Sevst’yanov. Branching stochastic processes for particles diffusing in a bounded domain with absorbing boundaries. *Theory of Probability and its Applications*, 1,3:111–126, 1958.
- [Tho06] V. Thomee. *Galerkin Finite Element Methods for Parabolic Problems*. Springer, Berlin, 2006.
- [Wat65] S. Watanabe. On the branching process for Brownian particles with an absorbing boundary. *Jour. Math. Kyoto Univ.*, 4:385–398, 1965.

[Wat67] S. Watanabe. *Limit theorem for a class of branching processes.*, pages 205–232. J. Chover ed., Markov Processes and Potential Theory. Wiley, New York, 1967.

DPMMS, UNIVERSITY OF CAMBRIDGE, CAMBRIDGE, CB3 0WA, ENGLAND.

E-mail address: ep361@cam.ac.uk